



Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Escuela Académico Profesional de Ingeniería de Sistemas

**Diseño de un datawarehouse para el Programa
Nacional de Apoyo Directo a los más Pobres - Juntos**

TESINA

Para optar el Título Profesional de Ingeniero de Sistemas

AUTORES

Miguel Santiago ÁVILA JARA
Patricia CHAMOCHUMBI ALCÁNTARA

ASESOR

Erwin MAC DOWALL REYNOSO

Lima, Perú

2007



Reconocimiento - No Comercial - Compartir Igual - Sin restricciones adicionales

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

Usted puede distribuir, remezclar, retocar, y crear a partir del documento original de modo no comercial, siempre y cuando se dé crédito al autor del documento y se licencien las nuevas creaciones bajo las mismas condiciones. No se permite aplicar términos legales o medidas tecnológicas que restrinjan legalmente a otros a hacer cualquier cosa que permita esta licencia.

Referencia bibliográfica

Ávila, M. (2007). *Diseño de un datawarehouse para el Programa Nacional de Apoyo Directo a los más Pobres - Juntos*. Tesina para optar el título de Ingeniero de Sistemas. Escuela Académico Profesional de Ingeniería de Sistemas, Facultad de Ingeniería de Sistemas e Informática, Universidad Nacional Mayor de San Marcos, Lima, Perú.

DEDICATORIAS

Quiero dedicar este trabajo a mis padres Jorge y Rosa, por su invaluable apoyo; a mi hermano Jorge Luis, el mejor tío del mundo; a mi esposo Oscar, comprensivo y encantador; y a mi hijo Said Gabriel, hermoso, inteligente y amoroso, mi mayor tesoro.

Patricia Chamocho Alcantara.

A mi madre que desde siempre se brinda por completo a sus hijos, a Dios por haberme permitido llegar a esta instancia de mi vida.

Miguel Ávila Jara.

INDICE

INDICE	3
INDICE DE IMÁGENES	10
RESUMEN	13
ABSTRACT	15
INTRODUCCIÓN	17
1. PLANTEAMIENTO DEL PROBLEMA	19
1.1. Fundamentación del problema	19
1.1.1. Descripción de la realidad	20
1.1.2. Antecedentes del Problema	30
1.2. Definición del Problema	33
1.3. Formulación del Problema	40
1.3.1. Objetivos Generales	40
1.3.2. Objetivos Específicos	40
1.4. Justificación de la investigación	40
1.4.1. Justificación de Carácter Práctico	40
1.4.2. Justificación de Carácter Metodológico	41
1.4.3. Justificación de carácter teórico	41
1.5. Finalidad e Importancia	41
1.5.1. Finalidad	41
1.5.2. Importancia	42
1.6. Limitaciones de la investigación	42
1.6.1. Delimitación temporal	42
1.6.2. Delimitación conceptual	42
1.6.3. Delimitación social	43
1.7. Alcance de la investigación	43
2. MARCO TEORICO CONCEPTUAL	44

2.1.	Antecedentes de la investigación	44
2.2.	Bases teóricas	46
2.2.1.	Conceptos de Inteligencia de Negocios	46
2.2.1.1	<i>Inteligencia de Negocios</i>	46
2.2.1.2	<i>Business Performance Managment (BPM)</i>	47
2.2.2.	Conceptos de Datawarehouse	51
2.2.2.1	<i>Estructura de un Datawarehouse</i>	51
2.2.2.2	<i>Metadata</i>	51
2.2.2.3	<i>Middleware</i>	52
2.2.2.4	<i>Mecanismos de Extracción</i>	52
2.2.2.5	<i>Mecanismos de Carga</i>	53
2.2.2.6	<i>Staging area.</i>	53
2.2.2.7	<i>Arquitectura de un Datawarehouse</i>	54
2.2.2.8	<i>Diseño de un Datawarehouse</i>	55
2.2.2.9	<i>Seguridad de un Datawarehouse</i>	56
2.2.2.10	<i>Calidad de datos en un Datawarehouse</i>	57
2.2.2.11	<i>Calidad de un Datawarehouse</i>	59
2.2.2.12	<i>Modelado del Datawarehouse</i>	60
2.2.2.13	<i>Esquema estrella</i>	60
2.2.2.14	<i>Esquema copo de nieve - SnowFlake</i>	62
2.2.2.15	<i>Esquema mixto</i>	63
2.2.2.16	<i>Factores de éxito en Datawarehouse</i>	64
2.2.2.17	<i>Uso del Datawarehouse</i>	65
2.2.3.	Bases Metodológicas para el diseño de un Datawarehouse	67
2.2.3.1.	<i>Metodología Inmon</i>	68
2.2.3.2.	<i>Metodología Kimball</i>	70
2.2.3.3.	<i>Metodologías propietarias</i>	72
2.2.4.	Metodologías ágiles para el desarrollo de aplicaciones de Inteligencia de Negocios	74
2.2.4.1.	<i>SCRUM</i>	74

2.2.4.2.	<i>Extreme Programming (XP)</i>	75
2.3.	Definición de Términos Básicos	76
2.3.1.	Identificación de hogares pobres	76
2.3.1.1	<i>Focalización</i>	76
2.3.1.2	<i>Focalización Geográfica</i>	77
2.3.1.3	<i>Focalización individual</i>	77
2.3.1.4	<i>Umbral de pobreza</i>	77
2.3.1.5	<i>Índice de pobreza</i>	78
2.3.1.6	<i>Proceso de Intervención</i>	78
2.3.1.7	<i>Filtración</i>	78
2.3.2.	Obtención de lista de hogares preseleccionados	79
2.3.2.1.	<i>Hogar Objetivo</i>	79
2.3.2.2.	<i>Hogar preseleccionado</i>	79
2.3.2.3.	<i>Titular del hogar</i>	79
2.3.2.4.	<i>Taller Institucional</i>	79
2.3.2.5.	<i>Taller comunal</i>	79
2.3.3.	Determinación de hogares validados en las asambleas comunales	80
2.3.3.1.	<i>Promotor Social</i>	80
2.3.3.2.	<i>Asamblea comunal para validación</i>	80
2.3.3.3.	<i>Hogar validado</i>	80
2.3.4.	Determinación de los hogares afiliados	81
2.3.4.1.	<i>Hogar Validado sin restricción RENIEC</i>	81
2.3.4.2.	<i>Hogar validado con restricción RENIEC</i>	81
2.3.4.3.	<i>Acuerdo de compromiso</i>	81
2.3.4.4.	<i>Hogar incorporado a los sectores</i>	81
2.3.4.5.	<i>Formato FJ-001</i>	81
2.3.4.6.	<i>Formato FJ.-002</i>	82
2.3.4.7.	<i>Formato FS-001</i>	82
2.3.4.8.	<i>Formato FE-001</i>	82
2.3.4.9.	<i>Institución educativa</i>	82

2.3.4.10.	<i>Establecimiento de salud</i>	82
2.3.4.11.	<i>Cambio de titular</i>	83
2.3.4.12.	<i>Homonimia</i>	83
2.3.4.13.	<i>Hogar afiliado</i>	83
2.3.4.14.	<i>Hogar desafiliado</i>	83
2.3.5.	Determinación de los representantes de los hogares beneficiados	84
2.3.5.1.	<i>Hogar Beneficiado</i>	84
2.3.5.2.	<i>Titular del Hogar Beneficiado</i>	84
2.3.5.3.	<i>Padrón Activo</i>	84
2.3.5.4.	<i>Padrón Pasivo</i>	84
2.3.5.5.	<i>Prestaciones directas a los hogares beneficiados</i>	84
2.3.6.	Control de condicionalidades	85
2.3.6.1.	<i>Condicionalidad</i>	85
2.3.6.2.	<i>Formato FE-002</i>	85
2.3.6.3.	<i>Formato FM-001</i>	85
2.3.6.4.	<i>Formato FS-002</i>	85
2.3.6.5.	<i>Formato FS-003</i>	86
2.3.6.6.	<i>Hogar suspendido</i>	86
2.3.6.7.	<i>Miembros de hogar sujetos al sector nutrición</i>	86
2.3.6.8.	<i>Miembros del hogar sujetos al sector salud</i>	86
2.3.6.9.	<i>Miembros del hogar sujetos al sector educación</i>	86
2.3.7.	En relación a la determinación de los representantes de hogares beneficiados sujetos a pago de los fondos condicionados	87
2.3.7.1.	<i>Incentivo económico-monetario</i>	87
2.3.7.2.	<i>Apertura de cuentas</i>	87
2.3.7.3.	<i>Abono</i>	87
2.3.7.4.	<i>Oficina bancaria</i>	87
2.3.7.5.	<i>Transportadora</i>	87
2.3.8.	En relación a la gestión de los procesos de la GOAL	88
2.3.8.1.	<i>Procesos</i>	88

2.3.8.2.	<i>Clientes</i>	88
2.3.8.3.	<i>Productos</i>	88
2.3.8.4.	<i>Proveedores</i>	88
3.	DESARROLLO DE SOLUCIÓN	89
3.1	Desarrollo del Plan	89
3.1.1.	Misión	89
3.1.2.	Análisis de riesgos	89
3.1.3.	Definición de roles para la administración del proyecto	91
3.1.4.	Equipo del Proyecto	93
3.1.5.	Recursos disponibles para la implementación del proyecto	94
3.1.5.1	<i>Documentación</i>	94
3.1.5.2	<i>Tecnología: Software</i>	94
3.1.5.3	<i>Tecnología: Hardware</i>	94
3.2	Identificación de Requerimientos	95
3.2.1.	Indicadores de Gestión para el proceso de Afiliaciones	95
3.2.1.1	<i>Subproceso de Carga de información de Censos</i>	95
3.2.1.2	<i>Subproceso de Validación por taller</i>	98
3.2.1.3	<i>Subproceso de Envío a Cotejo RENIEC</i>	98
3.2.1.4	<i>Subproceso de Firma de Convenio</i>	99
3.2.1.5	<i>Subproceso de Suspensiones Temporales</i>	99
3.2.1.6	<i>Subproceso de Exclusiones</i>	99
3.2.1.7	<i>Subproceso de Cambio de Titular</i>	100
3.2.1.8	<i>Información general del proceso de Afiliaciones</i>	100
3.2.2.	Indicadores de Gestión para el proceso de condicionalidades	101
3.2.2.1	<i>Subproceso de Incorporación a los Sectores</i>	101
3.2.2.2	<i>Subproceso de generación de formatos</i>	102
3.2.2.3	<i>Subproceso de evaluación de cumplimiento de compromisos</i>	104
3.2.2.4	<i>Información General del proceso de Condicionalidades</i>	105
3.2.2.5	<i>Información General de las instituciones</i>	105
3.2.3.	Indicadores de Gestión para el proceso de Pagos	106

3.2.3.1	Subproceso de Apertura de cuentas	106
3.2.3.2	Subproceso de Cierre de Padrón de Abonos	107
3.2.3.3	Subproceso de Puntos de Pago	108
3.2.3.4	Información General del proceso de Pagos	109
3.3	Modelamiento de Datos	110
3.3.1.	Modelo Estrella: Identificación de población censada	111
3.3.2.	Modelo Estrella: Identificación población censada pobre	112
3.3.3.	Modelo Estrella: Identificación población censada no pobre	113
3.3.4.	Modelo Estrella: Identificación de vivienda	114
3.3.5.	Modelo Estrella: Identificación del hogar afiliado	115
3.3.6.	Modelo Estrella: Identificación de los miembros objetivos	116
3.3.7.	Modelo Estrella: Identificación de Centro Educativos	117
3.3.8.	Modelo Estrella: Identificación de Establecimientos de Salud	118
3.3.9.	Modelo Estrella: Identificación de Hogares incorporados	119
3.3.10.	Modelo Estrella: Identificación de Promotores	120
3.3.11.	Modelo Estrella: Identificación de Hogares validados	121
3.3.12.	Modelo Estrella: Identificación de Hogares Evaluados	122
3.3.13.	Modelo Estrella: Identificación de hogares que incumplieron	123
3.3.14.	Modelo Estrella: Identificación de niños que incumplieron en Educación	124
3.3.15.	Modelo Estrella: Identificación de hogares que incumplieron en Salud	125
3.3.16.	Modelo Estrella: Identificación de Titulares abonados	126
3.3.17.	Modelo Estrella: Identificación de Cuentas bancarias aperturadas	127
3.3.18.	Modelo Estrella: Identificación de titulares del Padrón de Abonos	128
3.3.19.	Modelo Estrella: Identificación de Titulares suspendidos	129
3.3.20.	Modelo Estrella: Identificación de Errores	130
3.4	Arquitectura del Datawarehouse	131
3.5	Mapeo de Datos	133
3.6	Extracción de Datos y Transformación de datos	134
3.6.1.	Proceso ETL: Carga de información de censos	135
3.6.2.	Proceso ETL: Carga de información de condicionalidades	136

3.6.3.	Proceso ETL: Carga de información del padrón de abonos	137
3.6.4.	Proceso ETL: Carga de información histórica mensual	138
3.7	Carga del Datawarehouse	139
3.8	Liberación de la metadata	141
3.9	Plan de Prueba y validación del Datawarehouse	141
4.	CONCLUSIONES Y TRABAJOS FUTUROS	144
4.1	Conclusiones	144
4.2	Recomendaciones	145
4.3	Trabajos Futuros	146
	BIBLIOGRAFIA	147
	GLOSARIO	149
	ANEXO 1	151
	ANEXO 2	168
	ANEXO 3	169
	ANEXO 4	170

INDICE DE IMÁGENES

Figura 01: Mapa de Pobreza 2006 - Nivel departamental	21
Figura 02: Cantidad de distritos muy pobres con respecto al total por departamento	22
Figura 03: Ámbito de Trabajo del PNADP - Nivel departamental	23
Figura 04: Incorporación de distritos al PNADP por Etapa Censal	24
Figura 05: Hogares pobres y no pobres por Etapa Censal	25
Figura 06: Evolución de Hogares Beneficiados a Septiembre 2007	26
Figura 07: Evolución de Hogares Beneficiados a Septiembre 2007 por Departamento	27
Figura 08: Comparación entre Afiliados y Beneficiados a Septiembre 2007.....	28
Figura 09: Porcentaje de Beneficiados por Departamento de Setiembre 2007	29
Figura 10: Evolución en los primeros 4 meses del PNADP	30
Figura 11: Evolución de los procesos de Condicionalidades 2006	31
Figura 12: Motivos de No Afiliación.....	32
Figura 13: Datos históricos que se pueden obtener por proceso	34
Figura 14: Comportamiento del Servidor de base de datos	35
Figura 15: Horas de inversión de los analistas de información	36
Figura 16: Indicadores de Gestión por proceso	37
Figura 17: Ejemplo de indicador de Gestión para el proceso de Condicionalidades	38
Figura 18: Ejemplo de Dashboards Gerenciales del PNADP	45
Figura 19: Ejemplo de Reportes Gerenciales del PNADP.....	46
Figura 20: Arquitectura para una completa solución BPM.....	50
Figura 21: Diferentes Arquitecturas para un Datawarehouse.....	54
Figura 22: Topología de Calidad de un Datawarehouse	59
Figura 23: Esquema Estrella.....	62
Figura 24: Esquema Copo de nieve.....	63
Figura 25: Variables consideradas para medir éxito del Datawarehouse	65
Figura 26: Uso de un Datawarehouse	66
Figura 27: Percepción de los usuarios con respecto a la calidad de datos.....	66
Figura 28: Problemas encontrados en el uso de un Datawarehouse.....	67
Figura 29: Arquitectura de la metodología Inmon.....	69

Figura 30: Arquitectura de la metodología Kimball.....	71
Figura 31: Comparación de metodologías de vendedores de tecnología propietaria.	73
Figura 32: Riesgos y Amenazas frente a los objetivos de un Datawarehouse.....	90
Figura 33: Análisis de Riesgos.....	91
Figura 34: Estrategia de asignación de Roles	92
Figura 35: Equipo del proyecto	93
Figura 36: Modelo Estrella – Identificación de Población censada	111
Figura 37: Modelo Estrella – Identificación de Población censada pobre	112
Figura 38: Modelo Estrella – Identificación de Población censada pobre	113
Figura 39: Modelo Estrella – Identificación de Vivienda	114
Figura 40: Modelo Estrella – Identificación del hogar afiliado	115
Figura 41: Modelo Estrella – Identificación de los miembros objetivos	116
Figura 42: Modelo Estrella – Identificación de Centros Educativos.....	117
Figura 43: Modelo Estrella – Identificación de Establecimientos de salud.....	118
Figura 44: Modelo Estrella – Identificación de Hogares incorporados	119
Figura 45: Modelo Estrella – Identificación de Promotores	120
Figura 46: Modelo Estrella – Identificación de Hogares validados	121
Figura 47: Modelo Estrella – Identificación de Hogares evaluados.....	122
Figura 48: Modelo Estrella – Identificación de Hogares que incumplieron.....	123
Figura 49: Modelo Estrella – Identificación de miembros que incumplieron en Educación.....	124
Figura 50: Modelo Estrella – Identificación de hogares que incumplieron en Salud	125
Figura 51: Modelo Estrella – Identificación de titulares abonados	126
Figura 52: Modelo Estrella – Identificación de cuentas bancarias aperturadas	127
Figura 53: Modelo Estrella – Identificación de titulares del padrón de abonos	128
Figura 54: Modelo Estrella – Identificación de titulares suspendidos	129
Figura 55: Modelo Estrella – Identificación de Errores	130
Figura 56: Proceso ETL	132
Figura 57: Arquitectura del Datawarehouse	133
Figura 58: Proceso ETL - Carga de información de censos	135
Figura 59: Proceso ETL - Carga de información de condicionalidades.....	136

Figura 60: Proceso ETL - Carga de información del padrón de abonos	137
Figura 61: Proceso ETL - Carga de información histórica mensual	138
Figura 62: Parámetros y métricas de prueba.....	140
Figura 63: Parámetros y métricas de validación del Datawarehouse.....	142

RESUMEN

DISEÑO DE UN DATAWAREHOUSE PARA EL PROGRAMA NACIONAL DE APOYO DIRECTO A LOS MAS POBRES - JUNTOS

Nombres y Apellidos: Miguel Santiago Ávila Jara

Patricia Paola Chamochumbi Alcántara

Septiembre - 2007

Asesor : Erwin Mac Dowall Reynoso

Grado : Magíster

El presente trabajo tiene como objetivo el diseño de un Datawarehouse para el Programa Nacional de Apoyo directo a los más pobres – Juntos (PNADP), el cual recopila información mensual y detallada del resultado de los procesos core del PNADP.

El PNADP es un programa social del gobierno, dirigido a los hogares en situación de pobreza extrema, con la finalidad de entregarles un incentivo económico, siempre y cuando cumplan con los requisitos de afiliación y se comprometan a cumplir con los compromisos tendientes a mejorar la calidad de vida del hogar.

El objetivo del PNADP es contribuir a la disminución de la pobreza extrema del País. Por esta razón, los gestores del PNADP (responsables de las gerencias) necesitan analizar la información del resultado de los procesos core, en forma de indicadores de gestión (variables de medición de desempeño), los cuales permiten en primer lugar verificar que el PNADP cumple con las metas propuestas, así como también identificar los problemas y mejorar los procesos.

Actualmente, los indicadores de gestión son limitados. Si bien permiten medir el resultado de los procesos, no identifican los problemas. Por lo tanto, no se está aprovechando toda la información valiosa que se puede obtener de la base de datos. Además los gestores del PNDAP no tienen una herramienta que les permita analizar la información.

En este contexto, la finalidad de este trabajo es el diseño de un Datawarehouse flexible, que permita implementar indicadores de gestión útiles para una óptima medición de resultados y que además sirva de herramienta de análisis a los gestores, permitiéndoles tomar decisiones más efectivas hacia los objetivos del PNADP.

Palabras Claves:

- Datawarehouse
- PNADP
- Afiliación
- Indicadores de Gestión
- Gestores del PNADP

ABSTRACT

DESIGN OF A DATAWAREHOUSE TO “PROGRAMA NACIONAL DE APOYO DIRECTO A LOS MAS POBRES – JUNTOS”

Names and surnames: Miguel Santiago Ávila Jara

Patricia Paola Chamocho Alcantara

September - 2007

Asesor : Erwin Mac Dowall Reynoso

Grado : Magíster

This invest is about the design of a Datawarehouse to “Programa Nacional de Apoyo directo a los más pobres – Juntos (PNADP)”, which compiles monthly information about core process of PNADP.

PNADP is a government organization, which is focus in very poor people, in order to give them an economic incentive, as long as they fulfill requirements of affiliation and duties for improve their quality of life.

The purpose of PNADP is to contribute to reduce poverty in our country; so, managers need to analyze information on each core process using business indicators, which allow them cope with the main problems and optimize process.

There are not enough business indicators, making quite difficult identify the main problems. In consequence, not all valuable information in the database has been used. What is more, PNADP's managers have not any tool in order to analyze this information.

The purpose of this work is to develop a flexible Datawarehouse, which allows implement useful business indicators, which can enable an efficient measure of results. In addition, managers could use it as tool for an effective process of decision making, to reach organizational objectives.

Key words:

- Datawarehouse
- PNADP
- Afiliation
- Business indicators
- PNADP's managers

INTRODUCCIÓN

El presente trabajo consiste en el diseño de un Datawarehouse para el Programa Nacional de Apoyo Directo a los más pobres – Juntos, a quien en adelante se le citará como PNADP, el cual recopila información mensual y detallada de los principales procesos.

Actualmente el PNADP cuenta con bastante información, la cual no es aprovechada de manera óptima debido que el sistema de información es demasiado complicado de usar para los responsables en la toma de decisiones, además hay información que no se puede obtener desde el mismo sistema.

En septiembre del 2007, en el marco de su segundo aniversario, el PNADP se ha planteado cumplir nuevas metas, para lo cual se han identificado indicadores de gestión que permitirán analizar la situación general, así como la situación de cada región, distrito y hogar.

Por lo tanto, el presente trabajo es un aporte al PNADP, ya que es una solución de Datawarehouse que permitirá mostrar los indicadores de gestión necesarios para cumplir con los objetivos.

El primer capítulo comprende la fundamentación del problema, descripción de la realidad, antecedentes y definición del problema. Así mismo se identifican los objetivos, justificación, finalidad, importancia y alcance de la solución propuesta.

El segundo capítulo abarca los antecedentes de la investigación, conceptos de inteligencia de negocios, conceptos de Datawarehouse y la definición de términos básicos del PNADP.

El tercer capítulo comprende el desarrollo de la solución propuesta, en donde se identifican los requerimientos, los modelos diseñados en base a los requerimientos y los procesos de extracción, transformación y carga del Datawarehouse.

Finalmente, el cuarto capítulo enumera las conclusiones, recomendaciones y trabajos futuros.

CAPITULO I

1. PLANTEAMIENTO DEL PROBLEMA

1.1. Fundamentación del problema

En la actualidad, la cantidad información que maneja el PNADP es considerable, sólo en el mes de Agosto ha beneficiado a más de 300,000 hogares. Sin embargo, no hay herramientas que permitan analizar la información de manera efectiva.

Los gestores del PNADP consumen la mayor parte de su tiempo en obtener datos, en preparar cuadros estadísticos, etc. para entregar informes. Por esta razón, con una herramienta de Datawarehouse podrían invertir mejor su tiempo analizando la información, identificando los problemas y planteando estrategias de solución.

Para analizar la información los gestores utilizan indicadores de gestión, los cuales, aunque han ido evolucionando con el tiempo, siguen siendo simples. Esto se debe a que el sistema del que se obtiene la información es muy limitado.

Por lo tanto, la propuesta de solución es diseñar un Datawarehouse, con un modelo de datos flexible, que sirva de herramienta a los gestores permitiéndoles obtener la información de los indicadores de gestión de manera rápida y analizarla para cumplir más efectivamente a los objetivos del PNADP.

1.1.1. Descripción de la realidad

En la actualidad el Gobierno, a través del PNADP, está logrando tener un real acercamiento con aquella población muchas veces afectada por la pobreza y la violencia, asumiendo su responsabilidad a fin de lograr un desarrollo económico con justicia social, incentivándolos para que puedan mejorar su calidad de vida; así como obtener sus propios ingresos.

Para lograr este acercamiento, el PNADP utiliza instrumentos de focalización que sirven para identificar geográficamente los departamentos y distritos con mayor cantidad de población extremadamente pobre. Cada cierto tiempo los gestores seleccionan un conjunto de distritos para que la población pueda ser censada y beneficiarse del PNDAP.

La figura 01 es una muestra de instrumento de Focalización hecha por FONCODES en el 2006. El Mapa de Pobreza, clasifica los departamentos en 5 niveles de Pobreza. El nivel 1 agrupa los departamentos más pobres y el nivel 5 agrupa los departamentos menos pobres. Este estudio tomó como fuente de datos las características de la vivienda, y la carencia de los derechos básicos como agua, desagüe, electricidad, tasa de analfabetismo y cantidad de niños.



Figura 01: Mapa de Pobreza 2006 - Nivel departamental
Fuente: FONCODES

La figura 02 muestra la cantidad de distritos más pobres con respecto al total de distritos por departamento. Como se puede apreciar, el departamento de Cajamarca es el que tiene la mayor cantidad de distritos en extrema pobreza, seguido por Ayacucho y Huancavelica. También se puede apreciar que Huanuco es el departamento que tiene el mayor porcentaje de distritos en extrema pobreza con respecto al total por departamento.

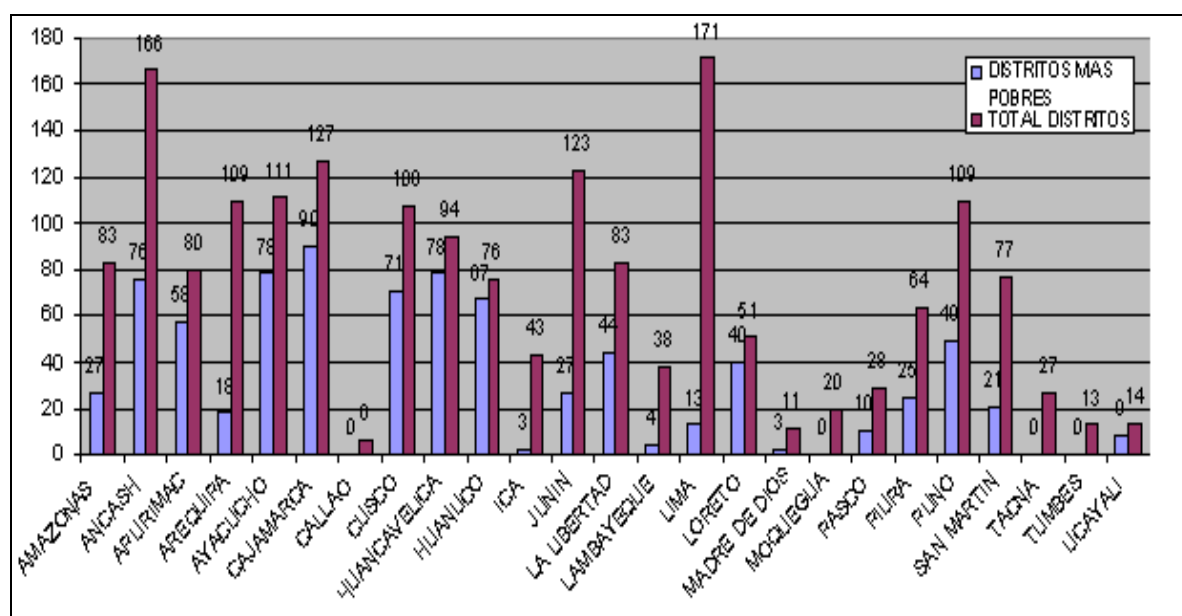


Figura 02: Cantidad de distritos muy pobres con respecto al total por departamento
Fuente: FONCODES

Gracias a los instrumentos de Focalización se puede priorizar la incorporación de los departamentos y distritos más pobres del Perú. La figura 03 muestra el ámbito de trabajo del PNADP. Los departamentos de Loreto, Piura, Amazonas, Pasco y Cusco se incorporaron a mediados del 2007.



Figura 03: Ámbito de Trabajo del PNADP - Nivel departamental
Fuente: GOAL-PNADP

El INEI, gracias a un convenio con el PNADP, se encarga de censar a los hogares de los distritos focalizados, en donde, se recopila la información necesaria de las características de la vivienda y de los miembros del hogar. Actualmente, el PNADP tiene en su cobertura 638 distritos. La figura 04 muestra el porcentaje de distritos incorporados por Etapa Censal (cada etapa agrupa varios censos)

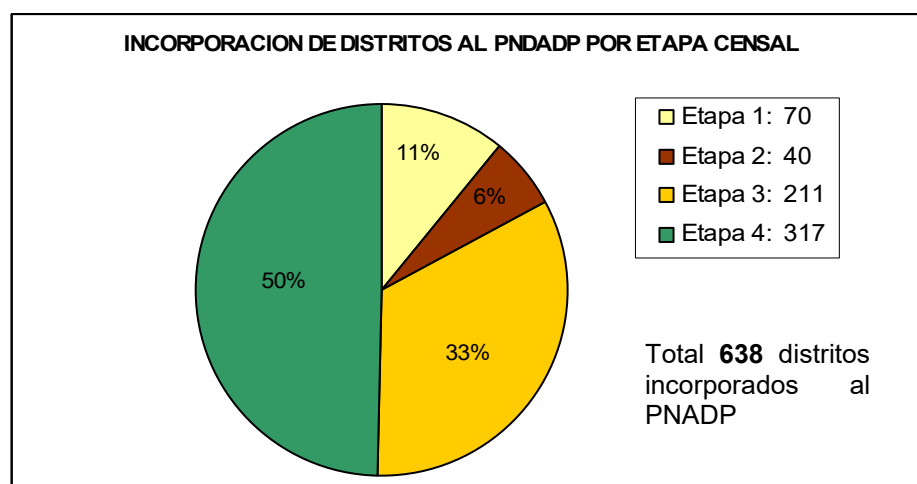


Figura 04: Incorporación de distritos al PNADP por Etapa Censal
Fuente: GTI – PNADP

La información de los hogares censados pasa por un proceso que determina el índice de pobreza del hogar. El resultado del proceso son dos listas: hogares pobres y no pobres respectivamente, las cuales son entregadas al PNADP para que se registren en la base de Datos. La finalidad es registrar toda la información que se recopiló del censo; sin embargo, solo los hogares Pobres se podrán afiliar al PNADP.

La figura 05 muestra la cantidad de hogares censados pobres y no pobres por Etapa Censal.

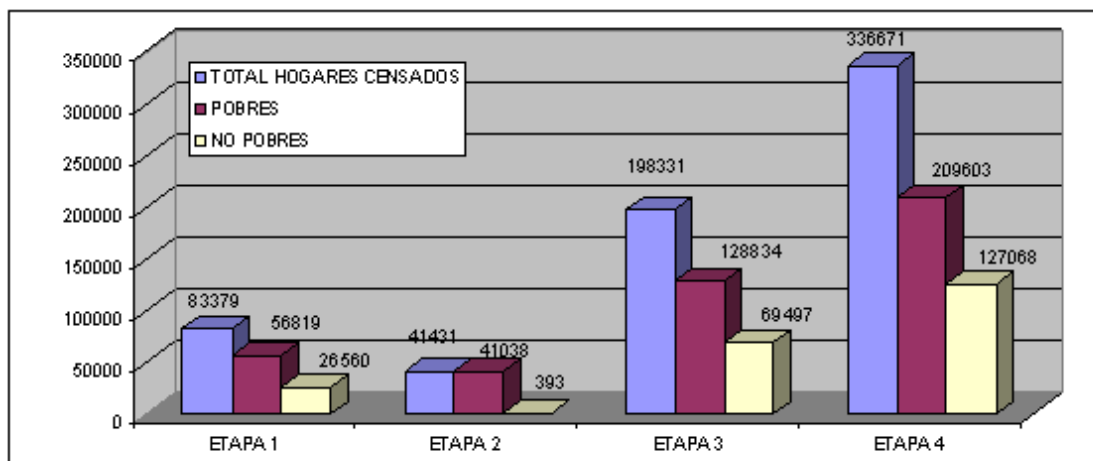


Figura 05: Hogares pobres y no pobres por Etapa Censal
Fuente: GTI – PNADP

El hogar pobre censado debe cumplir con los siguientes requisitos para ser beneficiado por el PNADP:

- Haber sido validado en una asamblea comunal que verifique que realmente el hogar sea pobre.
- La representante del hogar debe estar registrada en la RENIEC.
- Haber firmado el acuerdo de cumplimiento de compromisos.
- Tener una cuenta activa del Banco de la Nación.
- Cumplir con los compromisos de asistencia en los sectores Educación, Nutrición y Salud.

La figura 06, muestra la evolución del PNADP desde Septiembre del 2005 hasta Setiembre del 2007. Como se muestra en la línea azul, en Septiembre del 2005 benefició a 795 hogares. En Diciembre del 2005 benefició a 22,550 hogares. En Septiembre del 2007 benefició a 309,181 hogares. La línea naranja indica los hogares con validación comunal.

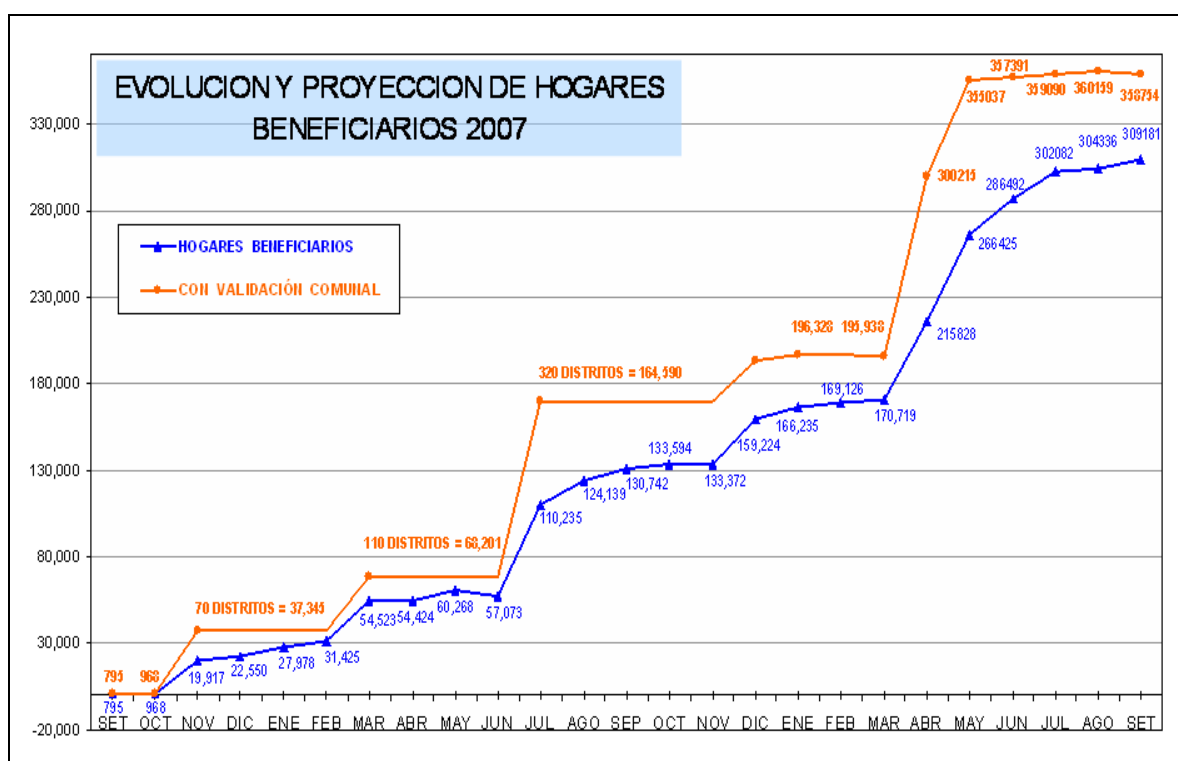


Figura 06: Evolución de Hogares Beneficiados a Septiembre 2007
Fuente: GOAL – PNADP

La figura 07, muestra la evolución del PNADP por Departamento. Se puede apreciar que en Septiembre del 2005 el único departamento beneficiado fue Ayacucho. Actualmente son 14 departamentos beneficiados.

La figura 08 muestra la cantidad de hogares afiliados y beneficiados al mes de Septiembre del 2007 por departamento. Se puede apreciar que Cajamarca es el departamento con mayor cantidad de hogares beneficiados y Amazonas el de menor cantidad de beneficiados.

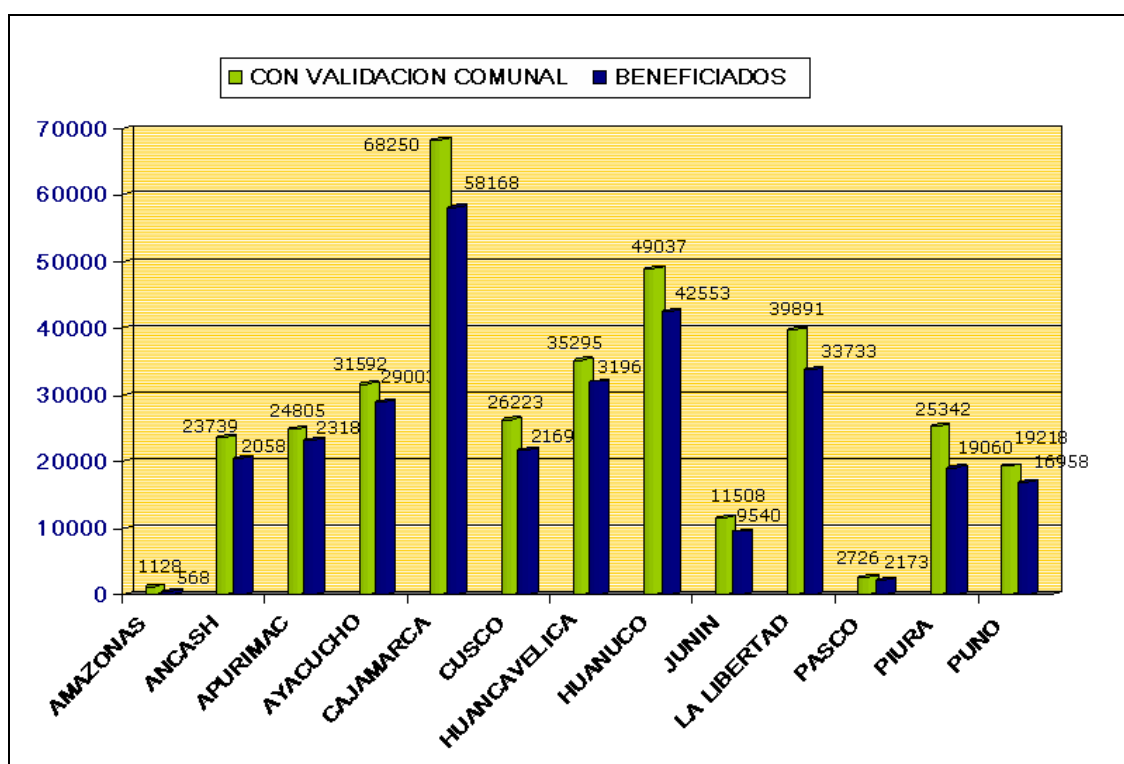


Figura 08: Comparación entre Afiliados y Beneficiados a Septiembre 2007
Fuente: GOAL – PNADP

La figura 09 muestra el porcentaje de beneficiados por departamento con respecto al total de beneficiados. Mientras que Amazonas representa el 0.18%, Cajamarca representa el 18.81 %.

DEPARTAMENTO	NRO DE HOGARES BENEFICIADOS	DISTRIBUCION %DEL TOTAL
AMAZONAS	568	0.18%
ANCASH	20580	6.70%
APURIMAC	23186	7.50%
AYACUCHO	29003	9.38%
CAJAMARCA	58168	18.81%
CUSCO	21697	7.02%
HUANCAVELICA	31962	10.34%
HUANUCO	42553	13.76%
JUNIN	9540	3.09%
LA LIBERTAD	33733	10.91%
PASCO	2173	0.70%
PIURA	19060	6.16%
PUNO	16958	5.48%
TOTAL	309181	100.00%

Figura 09: Porcentaje de Beneficiados por Departamento de Setiembre 2007
Fuente: GOAL – PNADP

1.1.2. Antecedentes del Problema

En Septiembre del 2005, el PNADP emitió la primera lista de hogares beneficiados, es decir, hogares que recibirían un incentivo económico de 100 soles. De aquí la meta para los próximos meses sería incrementar la cantidad de hogares beneficiados.

Para incrementar la cantidad de hogares beneficiados debían incrementar la cantidad de hogares censados y la cantidad de hogares afiliados. Por lo que estas variables se identificaron como indicadores de gestión para los procesos de Afiliaciones y Pagos. En el figura 10 se puede apreciar la evolución de los indicadores.

<i>Total de Hogares</i>	<i>2005</i>			
	<i>Septiembre</i>	<i>Octubre</i>	<i>Noviembre</i>	<i>Diciembre</i>
Censados	1048	1048	37345	37345
Afiliados	805	968	19975	22550
Beneficiados	805	968	19975	22550
No Afiliados	243	80	17370	14795

Figura 10: Evolución en los primeros 4 meses del PNADP
Fuente: GTI - PNADP

En Enero del 2006, se realizó el primer proceso de condicionalidades, el cual dio como resultado la primera lista de hogares que no cumplieron con los compromisos de educación, salud y nutrición; así como la primera lista de hogares suspendidos por no cumplir con alguno de los compromisos.

Dado que los hogares fueron suspendidos por un periodo de 3 meses, la cantidad de hogares beneficiados en ese período disminuyó, lo que creó una preocupación a los gestores. De aquí que la meta para los próximos procesos de Condicionalidades sería disminuir la cantidad de suspensiones.

Para disminuir la cantidad de suspensiones debían disminuir el porcentaje de incumplimiento de los hogares evaluados por cada sector. Por lo que estas variables se identificaron como indicadores de gestión para el proceso de condicionalidades. En la figura 11 se puede apreciar su evolución.

Total de Hogares	2006		
	MAR/ABR/MAY	JUN/JUL/AGO	OCT/NOV/DIC
<i>Evaluados</i>	51632	50183	128655
<i>Suspendidos</i>	6151 (12 %)	4377 (9%)	8887 (7%)
<i>Evaluados en Educación</i>	26465	26303	72442
<i>No cumplieron en Educación</i>	4266 (16%)	447 (2%)	836 (1%)
<i>Evaluados en Nutrición</i>	26099	26114	47207
<i>No cumplieron en Nutrición</i>	1300 (5%)	464 (2%)	809 (2%)
<i>Evaluados en Salud</i>	50463	49934	128568
<i>No cumplieron en Salud</i>	3698 (7%)	2213 (4%)	3246 (3%)

Figura 11: Evolución de los procesos de Condicionalidades 2006
Fuente: GTI – PNADP

Hasta Mayo del 2006, todos los procesos principales del PNADP los ejecutaba la gerencia de Tecnología de Información (GTI). Al final de cada proceso, GTI se encargaba de presentar a los gestores del PNADP la información de los indicadores de gestión de manera estadística, sólo cuando lo solicitaban se les entregaba la información al detalle.

A partir de Junio del 2006, los procesos fueron automatizados en un Sistema, denominado SISPPAC, por lo que GTI poco a poco empezó a delegar responsabilidades a los analistas de la gerencia de Operaciones, Afiliaciones y Liquidaciones (GOAL). Desde aquí, se empezaron a identificar nuevos indicadores de gestión, los cuales también se mostraban desde el sistema de manera estadística y detallada. En la figura 12 se puede apreciar los nuevos indicadores de gestión para analizar los hogares no Afiliados.

Total de Hogares No Afiliados	2007		
	Enero	Febrero	Marzo
Por no estar validado por RENIEC	21756	18959	17375
Por no haber firmado el Acuerdo de compromiso	3092	2590	1599

Figura 12: Motivos de No Afiliación
Fuente: GOAL- PNADP

1.2. Definición del Problema

Actualmente, los analistas de información del PNADP usan como fuente de datos el Sistema de Información y Soporte a los procesos de pagos, afiliaciones y condicionalidades (SISPPAC), el cual a través de consultas permite visualizar el resultado de los diferentes procesos.

Al cierre de cada proceso, los analistas exportan los datos desde el SISPPAC en un archivo Excel y lo importan a una herramienta estadística como SPSS, el cual les permite hacer consultas más simples y rápidas sobre los datos importados.

Desde el SPSS diseñan sus cuadros y gráficos estadísticos para cada uno de los indicadores de gestión, lo que les permite medir los resultados.

Por último, los indicadores de gestión son entregados a los responsables en la toma de decisiones (gestores) del PNADP.

En este contexto, el problema principal es que los analistas invierten mucho tiempo en diseñar indicadores de gestión; segundo, hay indicadores que no se pueden diseñar ya que son muy complejos; tercero, no hay una herramienta que permita visualizar los indicadores de una manera más amigable para los gestores del PNADP.

A continuación se enumeran los diferentes problemas que se originan por usar como fuente de datos un sistema transaccional.

- A. Hay procesos que se realizan constantemente, por lo tanto, si el analista quiere hacer una consulta para este tipo de procesos va a obtener un resultado que va a cambiar según el tiempo de búsqueda. Ejemplo: A las 10am había 1000 niñas de 5 años y a las 10:20am había 998 niñas de 5 años.

- B. El SISPPAC permite hacer consultas históricas para los procesos que se ejecutan por períodos pero no para los procesos que se realizan constantemente. Ejemplo: Puedo saber cuántos hogares no han firmado el acuerdo de compromisos a la fecha, pero no puedo saber cuántos hogares no habían firmado hace un mes. La figura 13 muestra el porcentaje de datos históricos que se puede obtener del Sistema por proceso del PNADP.

	<i>Proceso de Afiliaciones</i>	<i>Proceso de Condicionalidades</i>	<i>Proceso de Pagos</i>
<i>¿Que porcentaje de indicadores históricos actuales pueden obtenerse del Sistema?</i>	5%	60%	70%

Figura 13: Datos históricos que se pueden obtener por proceso
Fuente: GOAL- PNADP

C. El tiempo en obtener los resultados de las consultas, ya que no sólo se ha incrementado considerablemente la base de datos, sino que también ha crecido la cantidad de usuarios que utilizan el sistema, los que acceden para realizar consultas, actualizaciones y ejecutar procesos; generando de esta manera cuellos de botella en el servidor de base de datos. La figura 14 muestra el comportamiento del servidor de la base de Datos, en un escenario como el ya descrito.

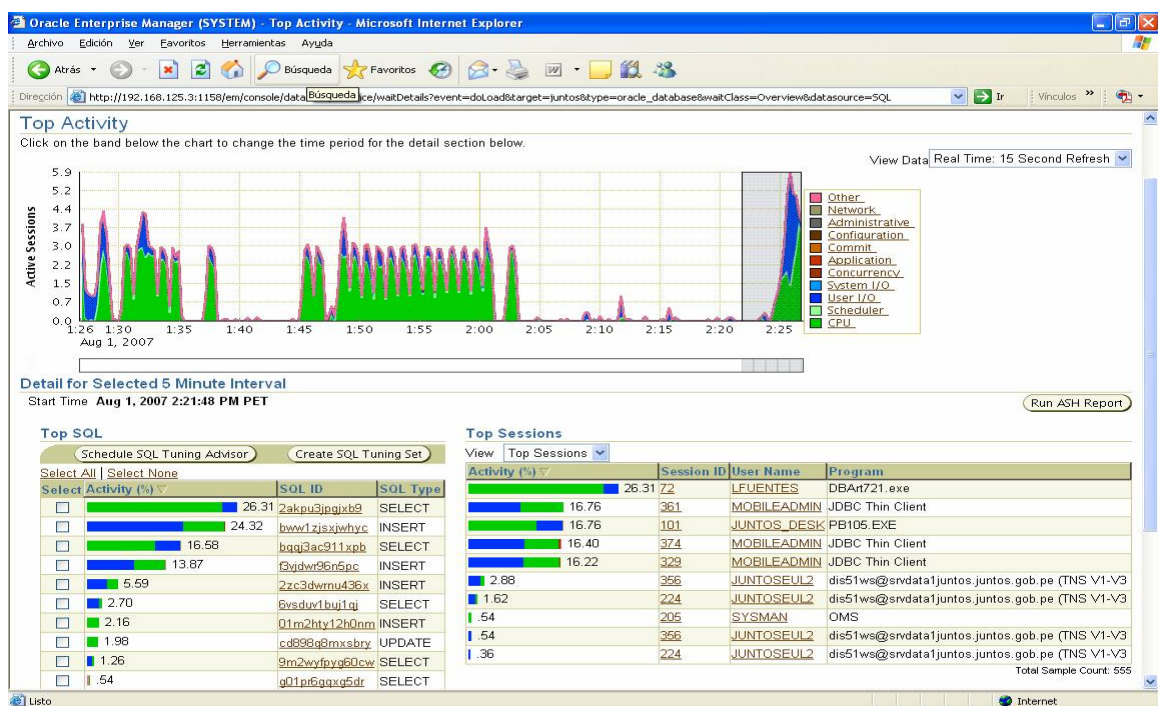


Figura 14: Comportamiento del Servidor de base de datos
Fuente: ORACLE Enterprise Manager, GTI – PNADP

D. Además de la demora en obtener el resultado nominal desde el SISPPAC, los analistas demoran en importar la información al SPSS y en diseñar cuadros y gráficos para cada indicador de gestión. La figura 15 muestra las horas que invierten los analistas para entregar el informe final de resultados a los gestores del PNADP.

Analista	Tiempo estimado (horas)	Tiempo estimado (días)	Cantidad de cuadros
Analista Afiliaciones	24	3	30
Analista Condicionalidades	31	4	26
Analista Pagos	15	2	20

Figura 15: Horas de inversión de los analistas de información
Fuente: GTI- PNADP

En la figura 16 se muestra los indicadores de gestión que se usan actualmente en los 3 procesos del PNADP y las horas en que aproximadamente demora en hacer los cuadros y gráficos por cada uno; tomando en cuenta que los indicadores deben presentarse agrupados por departamento y distrito.

hras	Analista de Afiliaciones	hras	Analista de Condiconalidades	hras	Analista de Pagos
2	Hogares Afiliados	2	Hogares no incorporados en Salud	1	Hogares no abonados por estar suspendido por condicionalidad
2	Hogares no afiliados	2	Hogares no incorporados en Educación	1	Hogares no abonados por no tener cuenta activa
2	Hogares que fueron suspendidos por Homonimia	1	Formatos generados para medir condicionalidades por sector	1	Hogares no abonados por no estar incorporado en el sector salud
2	Hogares que fueron excluidos del programa por no cumplir requisitos	4	Hogares y niños que debían ser evaluados por sector	1	Hogares no abonados por no estar incorporado en el sector educación.
1	Titulares que han fallecido	4	Hogares y niños que fueron evaluados por sector	1	Hogares abonados en modalidad Agencia
1	Titulares del Hogar que no tienen DNI	4	Hogares y niños que no cumplieron por sector	1	Hogares abonados en modalidad Transportadora
2	Titulares que han sido cambiados por otro miembro del hogar	1	Hogares fueron suspendidos	2	Distritos que tienen más de una agencia o transportadora.
2	Niños que no tienen fecha de nacimiento	2	Centros de salud que se les enviaron formatos	3	Hogares que fueron abonados el mes anterior y no el presente mes
2	Niños que no tienen partida de nacimiento	2	Instituciones educativas que se les enviaron formatos	1	Hogares que reciben su primer abono
2	Titulares que fueron enviados a cotejo de RENIEC	1	Formatos anulados por sector	3	Hogares que tenían más de un abono y recibieron este mes y no el mes anterior.
2	Titulares que fueron validados por RENIEC	1	Formatos que no fueron registrados por sector		
1	Hogares Censados	3	Hogares con más de dos suspensión por condicionalidad		
1	Hogares censados con inconsistencias	4	Evolución por Periodo y Sector		
1	Hogares censados Pobres				
1	Hogares censados No Pobres				
24		31		15	

Figura 16: Indicadores de Gestión por proceso
Fuente: GOAL- PNADP

La figura 17 muestra un ejemplo de cuadro a nivel de departamento para los indicadores de gestión del proceso de condicionalidades: Hogares Evaluados por Sector, hogares suspendidos por sector, además muestra el porcentaje de incumplimiento con respecto a los hogares evaluados.

DEPARTAMENTO	CATEGORIA	SECTOR		
		EDUCACION	NUTRICION	SALUD
ANCASH	HOG EVAL	3596	2830	5444
	HOG SUSP	104	215	184
	SUSP/EVAL	3%	8%	3%
APURIMAC	HOG EVAL	10520	9030	18563
	HOG SUSP	76	148	414
	SUSP/EVAL	1%	2%	2%
AYACUCHO	HOG EVAL	13851	10662	22588
	HOG SUSP	121	133	745
	SUSP/EVAL	1%	1%	3%
CAJAMARCA	HOG EVAL	5409	4667	9060
	HOG SUSP	26	80	68
	SUSP/EVAL	0%	2%	1%
HUANCAVELICA	HOG EVAL	13104	13531	26173
	HOG SUSP	77	146	735
	SUSP/EVAL	1%	1%	3%
HUANUCO	HOG EVAL	15525	3489	26916
	HOG SUSP	156	45	543
	SUSP/EVAL	1%	1%	2%
JUNIN	HOG EVAL	2036	0	3404
	HOG SUSP	38	0	78
	SUSP/EVAL	2%	-	2%
LA LIBERTAD	HOG EVAL	4489	0	9990
	HOG SUSP	189	0	285
	SUSP/EVAL	4%	-	3%
PUNO	HOG EVAL	3912	2998	6430
	HOG SUSP	44	39	150
	SUSP/EVAL	1%	1%	2%

Figura 17: Ejemplo de indicador de Gestión para el proceso de Condicionalidades
Fuente: Informe GOAL - PNADP

- E. Finalmente, el SISPPAC es limitado, hay indicadores que no se pueden construir porque la información no se muestra en el sistema. La información está en la base de datos pero no está implementada como consulta.

Para resolver estos problemas, es necesario facilitar al PNADP una solución de Datawarehouse, que almacene toda la información histórica para que de esta manera se pueda obtener todos los indicadores de gestión de una manera más ordenada y rápida.

Con una solución de Datawarehouse se puede diseñar una herramienta que permita visualizar los indicadores de gestión en forma de reportes nominales, cuadros y gráficos estadísticos, que sirva a los analistas y a los gestores para analizar la información y medir los resultados.

Por lo tanto, el tiempo que antes invertía el analista en obtener la información, construir los indicadores y diseñar los cuadros, lo puede utilizar en identificar los hogares con problemas, hacerles seguimiento y apoyarlos para que puedan superarlos y así los hogares podrán salir más rápido de la pobreza.

Así mismo, los gestores teniendo la información a su alcance en todo momento, podrán tomar mejores decisiones adelantándose a los problemas, así como mejorar los procesos de su gestión.

Finalmente, la solución propuesta va a permitir que se cumpla más efectivamente los objetivos del PNADP, de contribuir a la disminución de la pobreza del País.

1.3. Formulación del Problema

1.3.1. Objetivos Generales

Diseñar una solución de inteligencia de negocios para el PNADP. Incluirá un diseño de un *Datawarehouse* y un tablero de control

1.3.2. Objetivos Específicos

- a) Lograr una percepción integral de las necesidades de análisis de información del PNADP.
- b) Proporcionar las herramientas adecuadas para permitir a los gestores el análisis detallado de la información de los hogares beneficiados del PNADP.
- c) Implementar gráficamente los principales indicadores de negocio del PNADP.
- d) Implementar los procesos de administración adecuados para la sostenibilidad de la solución a implementar.

1.4. Justificación de la investigación

1.4.1. Justificación de Carácter Práctico

Es importante porque culminará con el diseño y la implementación de un Datawarehouse, el cual contribuirá a una mejora de la gestión del PNADP.

1.4.2. Justificación de Carácter Metodológico

Es importante porque pretende ser una guía, para la implementación de soluciones de Inteligencia de Negocios, en entidades sin fines de lucro, en el marco de la lucha contra la pobreza. El diseño de un Datawarehouse, sugerirá la aplicación de buenas prácticas que puedan ser replicadas en instituciones similares.

1.4.3. Justificación de carácter teórico

Es importante porque enriquecerá la concepción teórica, respecto a la aplicación de soluciones de inteligencia de negocio a entornos para los que típicamente no son aplicados.

1.5. Finalidad e Importancia

1.5.1. Finalidad

El presente trabajo pretende mostrar la aplicabilidad de la inteligencia de negocios, el cual es un concepto utilizado tradicionalmente en empresas con fines de lucro, en la lucha contra la pobreza en el Perú, en tal aspecto tomará como referencia al PNADP.

1.5.2. Importancia

Una solución de Inteligencia de Negocios es importante porque permitirá proporcionar herramientas a gestores (tableros de control), y a los analistas (análisis OLAP), para poder transformar los datos en conocimiento.

Permitirá evaluar el impacto de las acciones tomadas, pudiendo aplicar los correctivos de manera oportuna.

Ayudará a analizar la información a un nivel de detalle personalizado, de forma que puedan enfocarse en problemáticas particulares.

1.6. Limitaciones de la investigación

1.6.1. Delimitación temporal

Se realizara el estudio con data histórica del PNADP, comprendida desde su puesta en marcha, (Septiembre 2005), a la fecha.

1.6.2. Delimitación conceptual

Los conceptos comprendidos en la presente investigación abarcan Inteligencia de Negocios, específicamente Datawarehouse y análisis OLAP, tableros de mando e Indicadores de impacto social.

1.6.3. Delimitación social

Los involucrados directos son los gestores y analistas del negocio del PNADP. Los involucrados indirectos son: las instituciones o sectores con las que interactúa el PNADP para el cumplimiento de sus fines y los hogares beneficiados del PNADP.

1.7. Alcance de la investigación

La solución propuesta comprenderá hasta el diseño de los modelos de datos bajo la metodología propuesta; y la arquitectura conceptual del datawarehouse.

CAPITULO II

2. MARCO TEORICO CONCEPTUAL

En este capítulo se hará una revisión de los principales conceptos y metodologías involucradas en el desarrollo de soluciones de Inteligencia de negocios. Primero se mencionarán brevemente los antecedentes de la investigación, es decir, las soluciones de Inteligencia de Negocios que se han implementado en el PNADP. Luego, se explicarán los conceptos y algunas *tendencias en el campo de Datawarehouse. Finalmente se definirán los principales conceptos que se utilizan en el PNADP.*

2.1. Antecedentes de la investigación

En el PNADP se han desarrollado dos proyectos en el campo de inteligencia de negocios:

Dashboards Gerenciales

Utilizando el producto CORDA, se diseñaron *dashboards*, con algunos de los principales indicadores de negocio. Esta solución presenta algunos problemas, ya que la información es dinámica, es decir accede al entorno de producción para obtener los datos, similar a un sistema transaccional. Además, los indicadores mostrados, aunque permiten descender a nivel de distrito, no son personalizables, el analista no puede hacer un análisis ad-hoc.

La figura 18, muestra los principales *dashboards* diseñados en este proyecto.



Figura 18: Ejemplo de Dashboards Gerenciales del PNADP
Fuente: Portal Institucional PNADP

Reportes Gerenciales

Son reportes diseñados haciendo uso de la herramienta oracle Discoverer. Los reportes, hacen referencia a las necesidades particulares de una gerencia (Gerencia de Afilaciones y Liquidaciones - GOAL). Los datos se obtiene utilizando el mecanismo de vistas materializadas¹ con refrescamiento periódico sobre el entorno de producción.

¹ Se denomina vista materializada a la estructura de base de datos que se actualiza periódicamente en base a una consulta, en intervalos de tiempo definidos.

La figura 19 muestra el diseño de uno de los reportes elaborados en este proyecto. Un consolidado por ubigeo de hogares pobres, no pobres y no objetivos.

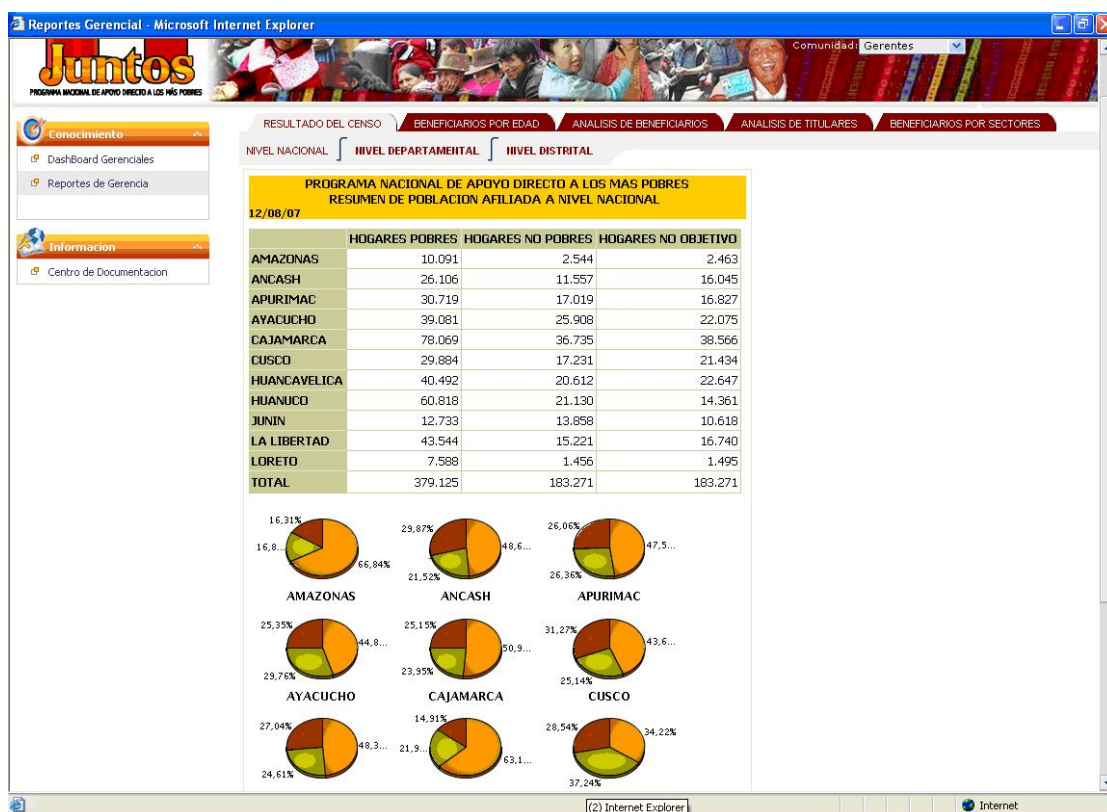


Figura 19: Ejemplo de Reportes Gerenciales del PNADP
Fuente: Portal Institucional PNADP

2.2. Bases teóricas

2.2.1. Conceptos de Inteligencia de Negocios

2.2.1.1 *Inteligencia de Negocios*

De acuerdo a [1], Inteligencia de Negocios es definida como “El proceso de transformar data en información y luego en conocimiento”.

“El conocimiento es típicamente obtenido sobre necesidades de los clientes, proceso de toma de decisiones de los clientes, la competencia, condiciones en la industria, y tendencias económicas, tecnológicas y culturales”.

De acuerdo a [5], “Las soluciones de Inteligencia de Negocios se basan en el Análisis avanzado de datos, por lo que requiere una infraestructura de Datawarehouse que puede soportarlo efectivamente “

2.2.1.2 Business Performance Managment (BPM)

Es un nuevo concepto de solución de inteligencia de negocios orientada a la optimización integral de la infraestructura del negocio. Por lo que una de prioridades es la investigación de los datos del negocio soportada en el Datawarehouse.

En [1] BPM es definido como *“un conjunto de procesos que ayuda a las organizaciones a optimizar el rendimiento del negocio, promoviendo la eficiencia de los procesos así como el uso eficiente de recursos humanos, financieros y materiales. BPM incluye Datawarehouse, pero también requiere un amplio conjunto de soluciones que descansan sobre diferentes tecnologías e impacta profundamente en toda la arquitectura de la plataforma de BI”.*

Las características que definen las diferencias entre BPM y el clásico datawarehousing son:

- *Usuarios, los usuarios de los sistemas BPM son aún tomadores de decisiones, pero a los niveles táctico y operacional. Estos usuarios tienen una limitada visión de la estrategia de la compañía, y solo tienen que ver con el subconjunto de indicadores relacionados a sus tareas específicas.*
- *Tiempo de entrega, las decisiones a los niveles más bajos deben ser más rápidas que las decisiones estratégicas.*
- *Nivel de detalle de la información y ciclo de vida. La información que circula normalmente en los sistemas BPM, es usualmente mas detallada que en los sistemas Datawarehouse. Sin embargo, el ciclo de vida de la información requerida por BPM es limitada, dado que los usuarios están interesados en el rendimiento actual de sus tareas.*
- *Interfase de usuario. Los que toman decisiones a nivel táctico y operacional seguramente no tendrán mucho tiempo o habilidades para hacer un análisis OLAP, así que la información será accedida principalmente en la forma de reportes y dashboards.*

La arquitectura para una completa solución BPM es mostrada en la figura 20. En el lado izquierdo de la figura se muestra la clásica arquitectura DW. En el lado derecho, la arquitectura se completa por un flujo de datos, mas orientado al monitoreo de procesos operacionales de tiempo critico. La tecnología que implementa este flujo es llamada monitoreo de la actividad del negocio – *Business Activity Monitoring* (BAM).

Los principales componente de BAM son:

- Un integrador - *Right-Time Integrator* (RTI).
- Un almacén de datos dinámico – *Dynamic Data Store* (DSS).
- Un administrador KPI.
- Un conjuntos de herramientas de minería de datos – *data mining*.
- Un motor de reglas – *rule engine*.

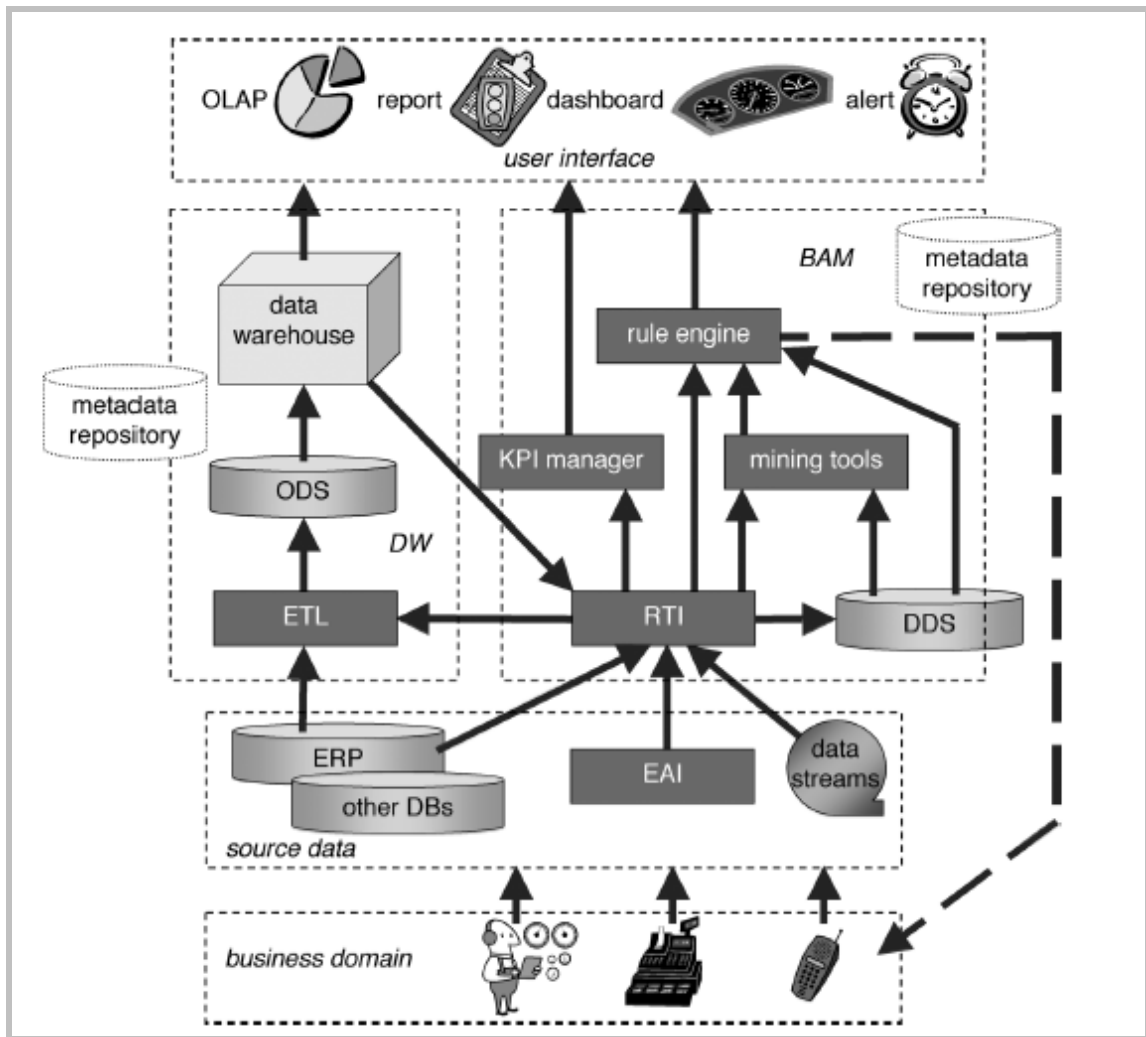


Figura 20: Arquitectura para una completa solución BPM
Fuente: [1]

Habiéndose revisado los conceptos y tendencias en el campo de Inteligencia de negocios, a continuación, se revisarán los aspectos técnicos del *Datawarehouse*. Los puntos a considerar son: estructura, arquitectura, diseño, factores críticos y tendencias en la implementación de un *Datawarehouse*.

2.2.2. Conceptos de Datawarehouse

2.2.2.1 Estructura de un Datawarehouse

Un Datawarehouse está compuesto por varios componentes sumamente importantes, entre los que se citan los siguientes:

- Metadata
- Middleware
- Mecanismos de Extracción
- Mecanismos de Carga
- Staging area

2.2.2.2 Metadata

En [12] se define la metadata como *"datos sobre los datos. Los metadatos constituyen un esquema que explica los datos que hay en el datawarehouse y donde están localizados. En un entorno, donde las relaciones entre los datos son tan importantes, el diccionario de datos define cómo se relacionan los datos entre sí"*.

Se diferencian los siguientes tipos de Metadata:

- Metadata de los sistemas fuente.
- Metadata del staging area.
- Metadata del manejador de base de datos.
- Metadata del "Front Room"
- Metadata del negocio.

2.2.2.3 Middleware

De acuerdo a [12] *"el middleware es el software que intercambia información en forma transparente entre aplicaciones y base de datos. El middleware ofrece un mecanismo de conexión abstracta entre el software de aplicación y la base de datos, y oculta al programador de aplicaciones de elementos específicos que dependen de la implementación"*.

Por lo tanto el Middleware tiene como función asegurar la conectividad entre los diversos componentes de la arquitectura de un datawarehouse.

2.2.2.4 Mecanismos de Extracción

Se considera la extracción de la data inicial o la modificada. Durante la carga inicial, capturar las modificaciones en la data origen resulta trivial, porque es probable que se realice la carga de la data total, pero una vez que esta carga se ha completado, la capacidad de capturar las modificaciones en la data origen cobra relevancia.

Determinar la estrategia apropiada para identificar los datos modificados puede ser un trabajo de profundo análisis.

Existen varias formas de lograrlo:

- Usar columnas de auditoria.
- Usar Logs.
- Usar extractos de tiempo.
- Procesos de eliminación.
- Cargas iniciales incrementales.

2.2.2.5 Mecanismos de Carga

Existen dos formas básicas de realizar el proceso de carga: Acumulación Simple y Rolling.

Acumulación Simple. Consiste en realizar una única sumarinización o resumen para todas las transacciones situadas en una ventana de tiempo determinada, de esta forma se transfiere al datawarehouse una única transacción.

Rolling. Se almacena la información de forma sumarinizada en diferentes niveles, generalmente de acuerdo a las agrupaciones de tiempo. Se aplica usualmente en los casos donde se requiere mantener varios niveles de granuralidad.

2.2.2.6 Staging area.

Puede verse el stage area como la “cocina” del Datawarehouse. Un área sólo accesible a personal altamente especializado en integración de datos de un *Datawarehouse*.

Usualmente se usa este término para describir los pasos discretos en el back room. *Staging* implica snapshots (vistas o fotografías rápidas) temporales o permanentes de la data.

2.2.2.7 Arquitectura de un Datawarehouse

La figura 21 muestra los principales tipos de arquitectura de un Datawarehouse:

- Arquitectura de un Datawarehouse empresarial.
- Arquitectura de un datamart.
- Arquitectura de un datamart hand hub spoke
- Arquitectura de un Datawarehouse con almacenamiento de data operacional.
- Arquitectura de un Datawarehouse distribuido.

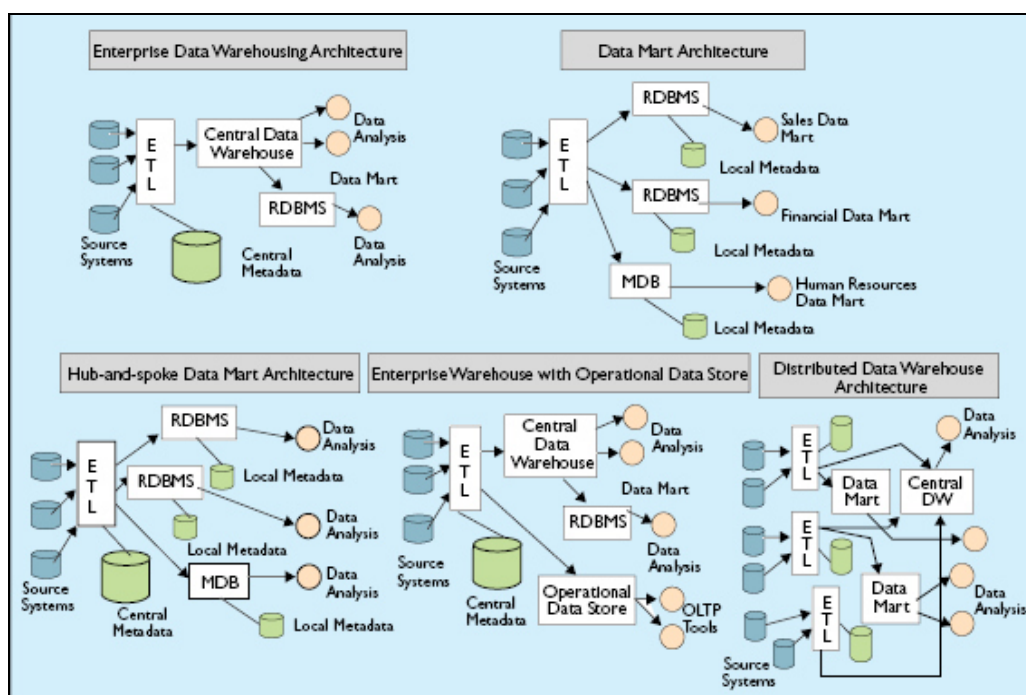


Figura 21: Diferentes Arquitecturas para un Datawarehouse
Fuente: [1]

2.2.2.8 Diseño de un Datawarehouse

De acuerdo a [4] *“hay 2 principales razones por las cuales es necesario la creación del Datawarehouse como un entorno de almacenamiento separado. La primera razón es que la performance de las consultas operacionales puede verse severamente afectada, si estas deben competir por los recursos computacionales con las consultas OLAP. La segunda razón tiene que ver con el hecho, incluso si la performance no sea el caso, de que es a menudo imposible estructurar una base de datos que pueda ser usada (consultada) de manera constante para propósitos operacionales y analíticos”.*

“La importancia de los Datawarehouses y el procesamiento analítico on-line (OLAP) como soporte para la toma de decisiones se ha incrementado rápidamente en los últimos años. Al mismo tiempo, un sentido de responsabilidad por la seguridad y privacidad ha surgido. Sin embargo, no se han hecho muchos estudios, para cubrir ambos campos”.

Es importante entender los factores que intervienen en una implementación exitosa de un *Datawarehouse*. Aspectos como seguridad, calidad de datos y calidad del *Datawarehouse*, son abordados con más detalle.

2.2.2.9 Seguridad de un Datawarehouse

El Datawarehouse por su naturaleza genera un conflicto de seguridad. Por otro lado, el objetivo es hacer que toda la información necesaria sea accesible de la forma más fácil posible. Así mismo, esta data es muy valiosa y sensible. El concepto de seguridad es bastante amplio (cubriendo aspectos éticos y sociales, morales, legales, etc.)

La comunicación ocupa un lugar crítico en un sistema Datawarehouse, creando la necesidad de adoptar mecanismos apropiados para una comunicación segura. El proceso de carga (transferencia de la data origen de las bases de datos operacionales al Datawarehouse) define nuevos requerimientos para la infraestructura de red. Bases de datos independientes (probablemente distribuidas), deben ser consolidadas sobre la red. Dado que la data puede ser altamente sensible, esta debe ser protegida, contra la interceptación y cualquier otra violación de seguridad.

En [3], se analizan los aspectos técnicos referidos a la seguridad y control de acceso.

Control de acceso

“El control de acceso en los extremos de la comunicación, involucra controlar el acceso al Datawarehouse y las bases de dato origen, mediante los procesos de extraer/cargar/transformar (ETL), y el acceso a esos procedimientos (así como su administración)”.

Diseño de seguridad OLAP

“La data de diferentes sistemas (con diferentes políticas) debe ser consolidada. Aunque los usuarios de los sistemas operacionales, no sean los mismos usuarios del Datawarehouse. Sin embargo el principal problema es, que el modelo relacional es el predominante en los sistemas operacionales, mientras que los sistemas OLAP hacen uso de los modelos multidimensionales. Los esquemas de control de acceso no interactúan fácilmente. La protección no es definida en términos de tablas, sino a nivel de dimensiones, granularidad, etc.”

2.2.2.10 Calidad de datos en un Datawarehouse

Los esfuerzos para la implementación de un Datawarehouse pueden fallar por varias razones. Primero es ignorar o trivializar los problemas con la data existente al inicio del proyecto o que surgen abruptamente durante el desarrollo del mismo.

La segunda razón es el fallo en el almacenamiento de la data apropiada. No es suficiente con marcar la data incorrecta o que no es útil, sino determinar qué es lo incorrecto con la data.

Aunque los aspectos relacionados a la calidad de los datos no son nuevos, lo que si es nuevo es el uso de la misma data para diferentes propósitos, los cuales pueden ser bastantes diferentes del original.

Como se indica en [9], *“Una de las principales características de los Datawarehouses es que estos son usados para la toma de decisiones más que para soporte a las operaciones. El soporte a varias decisiones, complica la administración de los datos, dado que probablemente se requieran varios niveles de calidad de datos”*.

La calidad de datos viene definida por las siguientes características [8]:

- La data es correcta, los valores y descripciones en la data describen sus objetos asociados verdadera y confiablemente.
- La data es no ambigua, se puede asumir que los valores y descripciones en la data tienen un sólo significado.

- La data es consistente, los valores y descripciones en la data usan una única constante, notación para su significado.
- La data es completa, es decir: los valores y descripciones individuales están definidos para cada instancia; y el número de registros “agregados” esta completo.

2.2.2.11 Calidad de un Datawarehouse

Esta referida a la calidad de la solución implementada. Esta viene definida por varias características, en [11] se incluye una topología con las principales características, la cual se muestra la figura 22.

Calidad del Sistema	
Atributos funcionales	Rendimiento / Flexibilidad / Confiabilidad / Extensibilidad / Mantenibilidad / Tiempos en actualizase / Seguridad
Atributos de usabilidad	Facilidad de uso / Facilidad de aprendizaje
Otras características	Captura de conocimiento / Transformación de datos / Integración de datos / Corrección de datos
Calidad de la información	
Atributos de entrada de datos	Relevancia / Disponibilidad / Cuantificabilidad / Exactitud / Multi-proposito / Integrabilidad / Agregabilidad
Atributos de salida de información	Claridad / Suficiencia / Trazabilidad / Valor Comparabilidad / Accesibilidad / Múltiples formatos de presentación

Figura 22: Topología de Calidad de un Datawarehouse
Fuente: [11]

2.2.2.12 Modelado del Datawarehouse

Los siguientes son los esquemas de modelado de datos más comunes de *Datawarehouse* [12].

2.2.2.13 Esquema estrella

“El esquema estrella es un paradigma de modelado que tiene un solo objeto en medio conectado con varios objetos de manera radial. El esquema estrella refleja la visión del usuario final de una consulta empresarial: hechos tales como ventas, pago y facturas se califican por una o más dimensiones (por mes, por producto, por región geográfica). El objeto en el centro de la estrella se denomina tabla de hechos y los objetos conectados a ella se denominan tablas de dimensión”.

“Un esquema estrella lógico sencillo consta de una tabla de hechos y varias tablas de dimensión. Los esquemas de estrella complejos, tienen cientos de tablas de hechos y de dimensión. Una tabla de hecho tiene las mediciones básicas de los negocios y consta de millones de registros. Las tablas de dimensión contienen atributos de negocios que se emplean como criterios de búsqueda SQL, y son relativamente pequeñas. Las siguientes son algunas de las técnicas que mejoran el desempeño de las consultas en el esquema estrella”:

- *“Definir adiciones en tablas de hechos existentes o nuevas tablas de adición. Por ejemplo, pueden existir ventas tanto de detalle como regionales en la misma tabla de hechos, con una misma columna indicadora de adición para diferenciar los registros. O bien, se puede crear una tabla de adición regional de ventas³”.*
- *“Segmentar la tabla de hechos de modo que la mayoría de las consultas solo accedan a un segmento o partición”.*
- *“Crear tablas de hechos separadas”.*
- *“Crear índices numéricos únicos u otras técnicas para mejorar el desempeño de las uniones”.*

En la figura 23 se muestra un diagrama con el esquema estrella. La tabla hecho es Invoice Item Facts, y las dimensiones son: Customer dimension, Location dimension, Product dimension y time dimension.

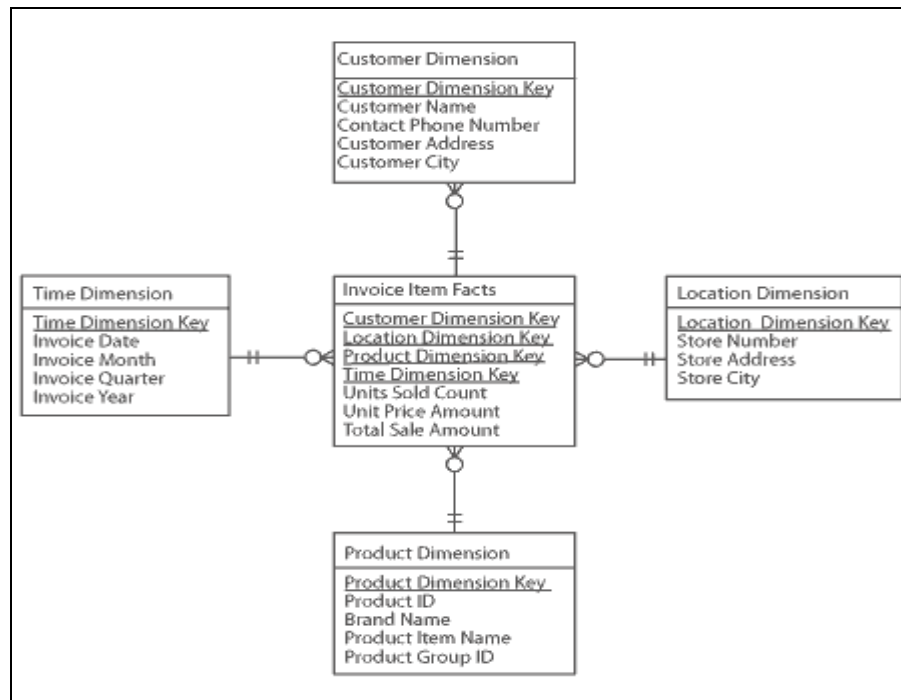


Figura 23: Esquema Estrella
Fuente: WWW.Datawarehouse.COM

2.2.2.14 Esquema copo de nieve - SnowFlake

“El esquema copo de nieve es una extensión del esquema estrella, en donde cada uno de los puntos de la estrella se divide en más puntos. En esta forma de esquema, las tablas de dimensión del esquema estrella contienen más normas. Las ventajas que proporciona el esquema de copo de nieve son mejorar el desempeño de consultas debido a un mínimo almacenamiento en disco para los datos y mejorar el desempeño mediante la unión de tablas más pequeñas con normas, en vez de las grandes tablas sin normas. Asimismo el esquema copo de nieve incrementa la flexibilidad de las aplicaciones debido a la aplicación de normas y, por lo tanto, disminuye la granularidad de las dimensiones”.

“El esquema copo de nieve aumenta el número de tablas con las que el usuario debe tratar e incrementa la complejidad de algunas consultas que debe localizar”.

En la figura 24 se muestra un diagrama con el copo de nieve. La tabla Product vendría a ser el hecho principal. Se puede observar que la dimensión región, es a su vez el hecho para las dimensiones Bundesland, PLZ-Gebiet y Stadt.

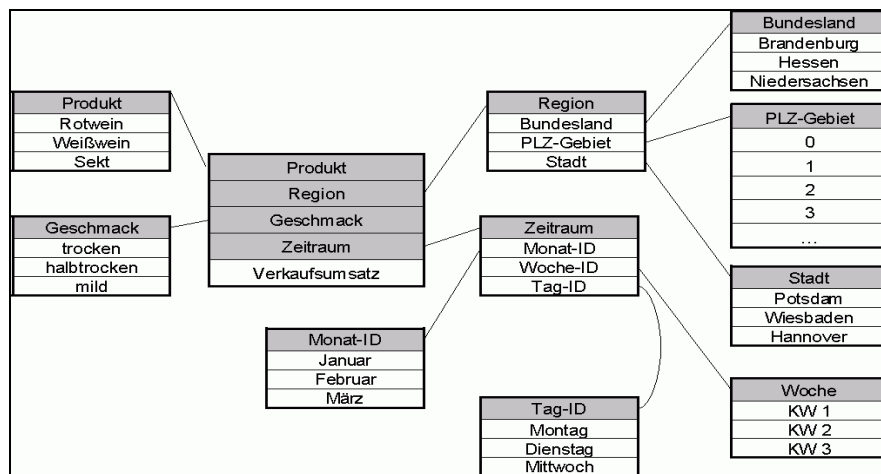


Figura 24: Esquema Copo de nieve
[Fuente: http://www.iwi.uni-hannover.de/lv/seminar_ws03_04]

2.2.2.15 Esquema mixto

“El esquema mixto es un convenio entre el esquema estrella, basado en tablas de hechos y tablas de dimensión sin normas, y el esquema de copo de nieve, en donde todas las tablas de dimensión tienen normas. En el esquema mixto, sólo se aplican normas a las tablas de dimensión más grandes. Por lo general, estas tablas contienen volúmenes de columnas sin norma (duplicados)”.

2.2.2.16 Factores de éxito en Datawarehouse

A pesar de su potencialidad, el éxito no está garantizado. De hecho, de acuerdo a [13], *“un estudio revela que más del 60% de las compañías que tienen Datawarehouse, indican que estos sólo tienen un éxito limitado en atender todas las expectativas de los usuarios”*. Según [14] *“Estudios también revelan que resistencia por parte del nivel administrativo al Datawarehouses por varias razones incluyendo el temor a perder el control, la falta de capacidades técnicas, costos e incertidumbre acerca de su utilidad”*.

A continuación se resumen el estudio realizado en [13], en el cual se analizan los principales factores para el éxito de un Datawarehouse desde diversas perspectivas. *“El modelo usado sugiere que hay seis principales dimensiones para el éxito de los sistemas de información: calidad del sistema, calidad de la información, calidad del servicio, uso, satisfacción del usuario y beneficios de la red (impactos individuales y organizacionales)”*

La figura 25 resume las variables utilizadas en dicho estudio.

Categoría de éxito	Variables estudiadas
Calidad de sistemas	Carga del sistema Facilidad de uso Facilidad para localizar la data Autorización de acceso
Calidad de datos	Relevancia Nivel de detalle Confiabilidad Consistencia
Calidad de la información	Utilidad de la información
Calidad del servicio	Entrenamiento al usuario
Satisfacción del usuario	Satisfacción del usuario

Figura 25: Variables consideradas para medir éxito del Datawarehouse
Fuente: [15]

El análisis de los resultados se basa en una escala del 1 al 7. Muy bajo a bajo (1-3), medio a medio alto (3.1-5), alto a muy alto (5.1 -7).

2.2.2.17 Uso del Datawarehouse

En general, la mayoría de los usuarios usan el Datawarehouse para labores no estructuradas, más que para labores rutinarias o administrativas.

La figura 26 muestra el resultado de un estudio, hecho en [13], sobre uso del Datawarehouse.

Tipo de Trabajo	Acceso	
	Si	No
Soporte a toma de decisiones	45 (70.3%)	19
Monitoreo del status	38 (59.4%)	26
Planeamiento	28 (43.7%)	36
Predicción	20 (31.2%)	44
Administración	19 (29.7%)	45
Contabilidad	19 (29.7%)	45
Localización de recursos/presupuesto	13 (20.3%)	51
Localización de recursos humanos/relaciones laborales	3 (7.8%)	59
Otros	12 (18.7%)	52

Figura 26: Uso de un Datawarehouse
Fuente: [13]

Con respecto a la percepción de los usuarios se puede decir que los aspectos más retadores en el Datawarehouse están en el campo de asegurar los adecuados niveles de consistencia. La figura 27 muestra el resumen de la percepción de la calidad de los datos por parte de los usuarios finales.

Calidad de datos	Promedio
Nivel de actualización de los datos	4.9 / 7
Nivel de detalle	4.9 / 7
Consistencia de los datos	4.1 / 7
Confiabilidad	4.6 / 7

Figura 27: Percepción de los usuarios con respecto a la calidad de datos
Fuente: [13]

Los principales problemas encontrados en este estudio están principalmente relacionados a la calidad de los sistemas y calidad de los servicios. Estos se muestran en la figura 28.

Calidad del sistema
Aspectos del diseño del sistema <ul style="list-style-type: none"> • Bajos tiempos de respuesta. • Muchos pasos para obtener la información. • Pobre modelado de datos. • Falta de pistas de auditoria. • Falta de metadata. • Accesibilidad • Limitados accesos al Datawarehouse • Aspectos de calidad de datos • Falta de exactitud de la data y datos perdidos. • Formato de datos anómalos • Registros duplicados • Nombres de campos inconsistentes • Baja confiabilidad de los datos
Calidad del servicio <ul style="list-style-type: none"> • Falta de un apropiado entrenamiento a los usuarios • Falta de información de contacto

Figura 28: Problemas encontrados en el uso de un Datawarehouse
Fuente: [13]

2.2.3. Bases Metodológicas para el diseño de un Datawarehouse

A continuación se exploran las principales metodologías para el diseño de un *Datawarehouse*.

2.2.3.1. Metodología Inmon

De acuerdo a [4], *“Esta opción concibe un Datawarehouse como una base de datos integrada, modelada usando la tradicional técnica de modelamiento entidad relación. Después que dicho Datawarehouse es creado, este sirve como origen de datos para los datamarts modelados dimensionalmente”*. La figura 29 muestra la arquitectura de esta opción.

La idea detrás de este método es tener un Datawarehouse almacenado de forma central, modelado con la técnica Entidad relación. Toda la integración de la data operacional, ocurre dentro del Datawarehouse centralizado. El proceso de integración y consolidación de la data, desde las base de datos operacionales en un Datawarehouse es conocido como proceso ETL. Desarrollar la infraestructura necesaria para los procesos ETL es usualmente la mas costosa parte (esfuerzo y tiempo) del diseño de un Datawarehouse [6]. Cuando un apropiado modelo es desarrollado, los requerimientos para el desarrollo de procesos ETL son claramente definidos.

Una vez que el Datawarehouse ha sido cargado con data, es posible obtener varios extractos para el análisis de datos. Uno de los tipos primarios de extractos analíticos, esta referido a los datamarts modelados dimensionalmente, los cuales son consultados mediante herramientas OLAP. El método Inmon reconoce la necesidad de otros, nodimensional, extractos de data potencialmente útil para el análisis y soporte a la toma de decisiones. Tales extractos no dimensionales, pueden incluir tablas, archivos planos, conjuntos de datos para datamining, etc. Inmon vio esos extractos, el Datawarehouse, y los repositorios de datos analíticos adicionales, como parte de un gran concepto llamado Fábrica de Información Corporativa - Corporate Information Factory (CIF). [7]

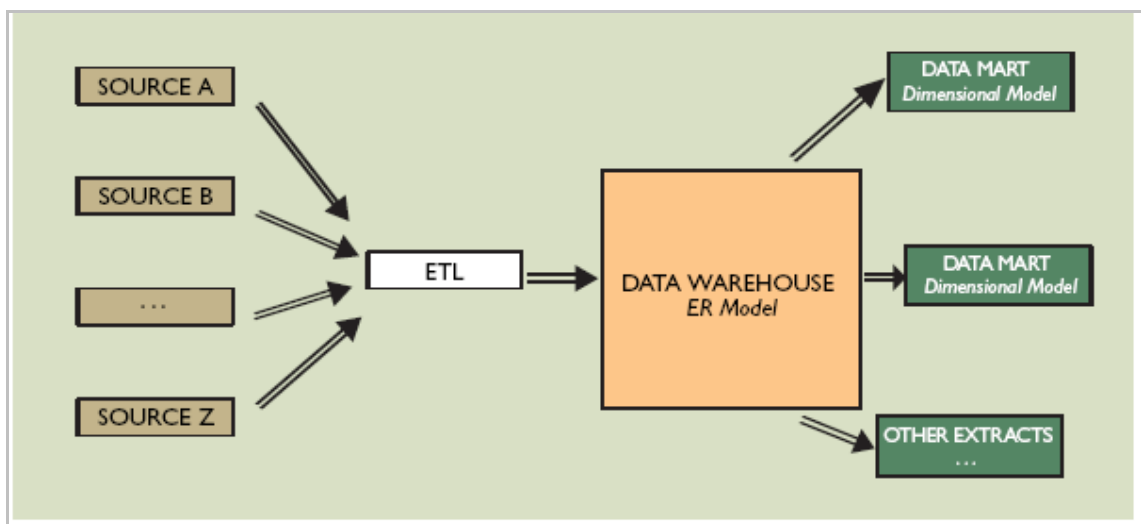


Figura 29: Arquitectura de la metodología Inmon
Fuente: [4]

2.2.3.2. Metodología Kimball

Concibe un Datawarehouse como una colección de datamarts modelados dimensionalmente. La figura 30 muestra la arquitectura de esta opción. Esta metodología es análoga a la metodología de Inmon, en lo referente a la utilización de las fuentes de datos operacionales y el proceso ETL. La diferencia es la técnica de modelado usada para modelar el Datawarehouse. Aquí, un conjunto de las dimensiones comúnmente usadas (por ejemplo tiempo), conocidas como “Dimensiones”, son primero identificadas. Las tablas de hechos – “Fact tables”, que corresponden a los sujetos de análisis, son luego añadidas. Un modelo dimensional es creado, donde cada Fact table es conectada a múltiples dimensiones, y algunas de las dimensiones son compartidas por mas de una Fact table. Adicionalmente, al conjunto de dimensiones originalmente creadas, algunas dimensiones adicionales pueden necesitarse. El resultado es un Datawarehouse que es una colección de datamarts modelados dimensionalmente e interconectados.

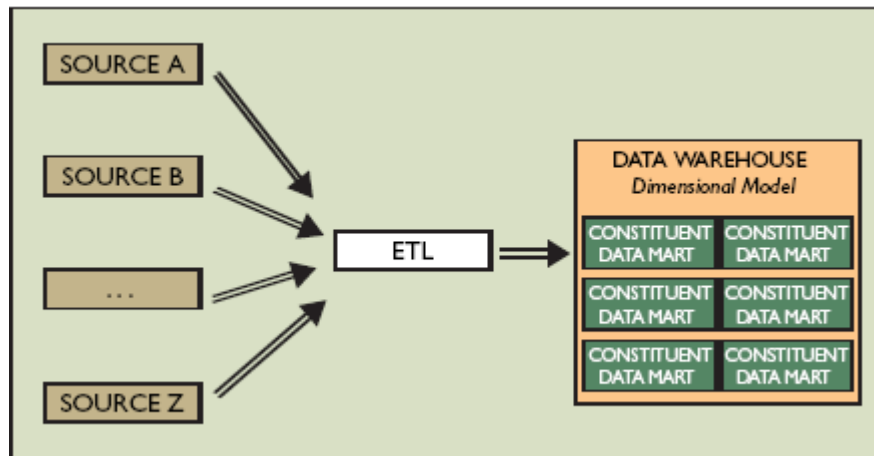


Figura 30: Arquitectura de la metodología Kimball
Fuente: [4]

Las ventajas y desventajas de ambas metodologías pueden ser, de acuerdo a [4] *“como la comparación entre la extensividad y el poder, versus la rapidez y la simplicidad. La metodología de Inmon requiere la creación de un Datawarehouse, modelado con ER, primero. El resultado de este proceso, puede ser luego utilizado para los siguientes pasos y como base para los extractos dimensionales y nodimensionales. En la metodología de Kimball, las estructuras modeladas dimensionalmente, son creadas sin necesidad de crear un modelo ER para estas. Si las estructuras dimensionales, son todo lo que la organización requerirá para cubrir sus necesidades de análisis, la metodología Kimball es la más rápida y fácil manera de crear un Datawarehouse. Sin embargo, si otros tipos de orígenes de datos analíticos son necesarios, adicionalmente a las estructuras dimensionales, la metodología de Inmon sería la más apropiada”*.

Las organizaciones deben elegir entre ambas metodologías en base a su naturaleza, a las necesidades actuales y futuras de análisis. Esto debería indicar cual método es el más apropiado para su entorno.

2.2.3.3. Metodologías propietarias

De acuerdo a [10], “la administración de los cambios es un punto importante a considerar en la selección de la metodología del Datawarehouse. Sorprendentemente, muy pocos vendedores incorporan la administración de los cambios en sus metodologías”.

La figura 31 muestra el resultado de la comparación entre las principales metodologías de vendedores de tecnología propietaria para el diseño de un Datawarehouse. Los vendedores que se han tomado en cuenta son: NCR, ORACLE, IBM DB2, SYBASE, SQL SERVER. Las características que se van a comparar son: competencia core, requerimientos de modelado, modelado de datos, soporte para normalización, diseño de arquitectural, estrategia de implementación, administración de metadata y escalabilidad.

	Metodologías Propietarias				
Atributos	NCR	ORACLE	IBM DB2	Sybase	SQL Server
Competencia Core	Teradata DBMS	ORACLE DBMS	DB2 DBMS	Sybase DBMS	SQL Server DBMS
Requerimientos de modelado	Entrevistas, priorización, plantillas, documentos de análisis	Entrevistas, priorización, subject areas.	Entrevistas.	Entrevistas	Documentos de análisis de entrevistas
Modelado de datos	ERD, esquema relacional.	Modelo dimensional, esquema estrella.	Modelo dimensional, esquema estrella.	ERD, esquema estrella, esquema relacional	Modelo dimensional, esquemas estrellas y snowflake.
Soporte para normalización / desnormalización	Desarrolla todas las relaciones normalizadas, permite normalizar.	Permite ambos.	Permite ambos.	Mas inclinado hacia la desnormalización.	Permite ambos.
Diseño de Arquitectura	Enterprise data warehouse	datamarts	Enterprise Datawarehouse y datamart	datamarts	Enterprise Datawarehouse y datamart
Estrategia de implementación	Iterativa	Ciclo de vida dimensional	Iterativa (prototipos)	Iterativa (RAD)	Iterativa
Administración de metadata	Si, usa un repositorio	Si, usa un repositorio	Si, usa un repositorio	Si, usa un repositorio	Si, usa un repositorio
Escalabilidad	Si, a cientos de terabytes.	No reportado	si	No reportado	Si, a cientos de terabytes
Administración de cambios	Ha publicado estudios, pero no se enfatiza en la metodología	No reportado		Considerado en la metodología	No reportado

Figura 31: Comparación de metodologías de vendedores de tecnología propietaria.
Fuente: [10]

El desarrollo de las aplicaciones de Inteligencia de negocios, entiéndase reportes, tableros de control o *dahsboards*, implica involucrarse en un proyecto de desarrollo, que dependiendo del contexto, puede abordarse como un proyecto aparte. A continuación se analizan las metodologías ágiles, los cuales han cobrado importancia en los últimos años como método para el desarrollo de las aplicaciones de Inteligencia de negocios.

2.2.4. Metodologías ágiles para el desarrollo de aplicaciones de Inteligencia de Negocios

2.2.4.1. SCRUM

“Scrum es una metodología para el desarrollo ágil de productos, expuesta por Hirotaka Takeuchi e Ikujiro Nonaka, en el artículo The New Product Development Game [Harvard Business Review Ene-Feb 1986]”, en el que ponen de manifiesto que:

- El mercado competitivo de los productos tecnológicos, además de los conceptos básicos de calidad, coste y diferenciación, exige también rapidez y flexibilidad.
- Los nuevos productos representan cada vez un porcentaje más importante en el volumen de negocio de las empresas.

2.2.4.2. Extreme Programming (XP)

El concepto de Extreme programming fue definido por Kent Beck. Recientemente esta metodología ha ganado popularidad dentro de la comunidad de la programación orientada a objetos. XP enfatiza el trabajo en equipo y el involucramiento de los usuarios. La comunicación y la retroalimentación son aspectos claves en XP.

Practicas clave en XP

- Programación en pares. Es la práctica de dos programadores o ingenieros ubicados contiguamente y participando en el desarrollo de una unidad de programación. Generalmente cada miembro desarrolla una tarea que el otro no esta haciendo. La proximidad y el desarrollo de una unidad de programación promueve el compartimiento de ideas y un mejor diseño.
- Refactoring. Es la actividad de continuo rediseño de una unidad de programación para aprovechar las ventajas de las diversas técnicas de programación, en especial de las orientadas a objetos y diseño de patrones, obteniendo programas mas reusables, simples y eficientes.
- Testing. Incluye las ideas de unidades de testeo, testeos funcionales (realizados por los usuarios), y

testeos automatizados. Estos conceptos soportan al concepto de unidades de programación y promueven la entrega de unidades libres de errores.

- Integración continua. Es el concepto de integrar nuevo código en el código existente, para luego utilizar las técnicas de testeo definidas en XP.
- Diseño evolutivo. Implica hacer iteraciones de los programas en minutos, en vez de días. Los pares de programación definen cada iteración del problema y luego las implementan.

2.3. Definición de Términos Básicos

En esta sección se definirá los conceptos básicos usados en el PNADP.

2.3.1. Identificación de hogares pobres

2.3.1.1 Focalización

Es un proceso en la administración pública orientada a la gestión del financiamiento de los recursos públicos a los Programas y Proyectos de naturaleza social empleando diversos instrumentos y metodologías a cargo de las entidades competentes (MEF, INEI); constituye un elemento fundamental tanto para la asignación eficiente del gasto público así como, durante la ejecución misma de dicho gasto.

Conlleva a la identificación de la población en situación de pobreza o de imposibilidad de acceso a determinados bienes o servicios indispensables para mantener niveles de vida aceptables.

2.3.1.2 Focalización Geográfica

Proceso que tiene por objetivo identificar áreas geográficas en las cuales se concentran la población con problemas de pobreza, vulnerabilidad social predominante que se desean resolver, estableciendo un ordenamiento geográfico para priorizar la asignación del gasto público.

2.3.1.3 Focalización individual

Tiene por finalidad identificar individuos u hogares en situación de pobreza por medio de indicadores elaborados con información individual. Este tipo de instrumento reconoce la heterogeneidad al interior de los agregados geográficos.

El hogar constituye la unidad de análisis para efectuar las estimaciones de pobreza, su definición permite definir su composición en base a ciertos criterios, y establece la pertenencia del individuo en el hogar así como sus miembros.

2.3.1.4 Umbral de pobreza

Un valor numérico que sirve de instrumento para identificar si el hogar (por su índice de pobreza) es pobre.

2.3.1.5 Índice de pobreza

Un valor numérico, que se obtiene como resultado de la aplicación del algoritmo empleado por INEI al procesamiento de las fichas socioeconómicas del barrido censal. Este valor se compara con el valor de otro índice denominado “Umbral de pobreza” para calificar la condición del hogar como pobre o no pobre

2.3.1.6 Proceso de Intervención

Proceso de Implementación del PNADP en un determinado espacio territorial y en un determinado tiempo, actividad que se desarrolla de manera participativa con las autoridades regionales, provinciales, distritales, las autoridades comunales y la propia población a ser beneficiada.

2.3.1.7 Filtración

Hogares afiliados al PNADP, sin tener las condiciones del hogar objetivo por su situación socio-económica y/o tener entre sus miembros niños de 0 a 14 años y madres gestantes. Es considerado un tipo de error por la falta de precisión en la búsqueda del grupo objetivo que el Programa ha definido y que resulta del empleo de los instrumentos de focalización geográfica e individual.

2.3.2. Obtención de lista de hogares preseleccionados

2.3.2.1. Hogar Objetivo

Está referido a los hogares en cuyos miembros se encuentran niñas y niños entre 0 a 14 años de edad y/o madres gestantes.

2.3.2.2. Hogar preseleccionado

Hogar objetivo que resulta de la identificación de su condición de extrema pobreza en el proceso del barrido censal, lo que constituye un hogar potencial a ser afiliado al PNADP.

2.3.2.3. Titular del hogar

Es el resultado de un proceso de selección del representante del hogar, de acuerdo a un criterio de prioridad según parentesco y sexo.

2.3.2.4. Taller Institucional

Evento en el cual se difunde los fines del PNADP a nivel de distrito; dirigido a autoridades locales y comunales representantes de instituciones públicas y privadas.

2.3.2.5. Taller comunal

Evento en el cual se difunde los fines y objetivos del PNADP a nivel de centros poblados; dirigido a la población en general.

2.3.3. Determinación de hogares validados en las asambleas comunales

2.3.3.1. Promotor Social

Colaborador del PNADP, asignado a uno o más distritos, incluidas comunidades y anexos. Tienen a su cargo la atención de un grupo de representantes del hogar beneficiado. Reporta directamente al Coordinador Regional respectivo y su labor administrativa depende de las directivas emanadas por la Gerencia de Coordinación Regional.

Constituye la extensión operativa de la gerencia de Operaciones, Afiliaciones y Liquidaciones, con la que coordinan la ejecución de los planes de expansión y el registro, recojo y monitoreo del cumplimiento de las condicionalidades por parte de los titulares de hogar acreditados. Es quien representa al PNADP en las distintas comunidades

2.3.3.2. Asamblea comunal para validación

Reunión de participación comunal que tiene como finalidad legitimar la incorporación del hogar preseleccionado por su condición de pobreza.

2.3.3.3. Hogar validado

Hogar preseleccionado y legitimado por la Asamblea de Validación Comunal

2.3.4. Determinación de los hogares afiliados

2.3.4.1. Hogar Validado sin restricción RENIEC

Hogar sin observación en el proceso de cotejo realizado por RENIEC.

2.3.4.2. Hogar validado con restricción RENIEC

Hogar con observación en el proceso de cotejo, realizado por RENIEC.

2.3.4.3. Acuerdo de compromiso

Documento que formaliza la incorporación del hogar validado al PNADP, estipulando las obligaciones de ambas partes.

2.3.4.4. Hogar incorporado a los sectores

Hogar cuyos miembros sujetos a evaluación de compromisos se encuentran formalmente registrados en el establecimiento de Salud y/o Institución Educativa, elegido por el representante del hogar.

2.3.4.5. Formato FJ-001

Documento de acuse de incorporación, en el cual se registran los datos del representante o titular del hogar validado.

2.3.4.6.Formato FJ.-002

Documento de identificación provisional, en el cual se registran los datos del representante o titular del hogar validado.

2.3.4.7.Formato FS-001

Documento de identificación del centro de salud, en el cual se registra la inscripción de los miembros objetivos a los establecimientos de salud.

2.3.4.8.Formato FE-001

Documento de identificación de la institución educativa, en el cual se registra la inscripción de los miembros objetivos (niños y niñas de 6 a 14 años) a las Instituciones Educativas.

2.3.4.9.Institución educativa

Centro de enseñanza de todo nivel cuyo ubigeo se encuentra dentro de la cobertura geográfica del PNADP.

2.3.4.10. Establecimiento de salud

Unidad operativa de la oferta de servicios de salud, clasificado en una categoría e implementada con recursos humanos, materiales y equipos encargada de realizar actividades asistenciales y administrativas que permiten brindar atenciones sanitarias tanto intramural como extramural, de acuerdo a su capacidad resolutive y nivel de complejidad.

2.3.4.11. Cambio de titular

Proceso autorizado y realizado por los colaboradores de la plataforma de GOAL al requerimiento que efectúa cada Coordinador responsable de las sedes regionales mediante documento o comunicación cierta, la que es evaluada para determinar su factibilidad.

2.3.4.12. Homonimia

Es un problema actual que enfrenta el PNADP, ya que se han dado casos de hacer abonos a personas homónimas. Cuando se identifica el problema se tramita con el Banco de la Nación para la devolución del dinero.

2.3.4.13. Hogar afiliado

Es el hogar validado sin restricción RENIEC y con acuerdo de compromiso firmado e incorporado a los sectores

2.3.4.14. Hogar desafiliado

Operación mediante el cual, una persona deja de formar parte del padrón de afiliados, por haber hecho de cambio de titular.

2.3.5. Determinación de los representantes de los hogares beneficiados

2.3.5.1. Hogar Beneficiado

Hogar afiliado que tenga por lo menos un abono.

2.3.5.2. Titular del Hogar Beneficiado

Es el representante del hogar beneficiado.

2.3.5.3. Padrón Activo

Registro nominal (nombres, apellidos y DNI) constituido por los titulares de los hogares afiliados sujetos a recibir el incentivo económico en el presente mes.

2.3.5.4. Padrón Pasivo

Registro nominal (nombres, apellidos y DNI) constituido por los titulares de los hogares afiliados que no van a recibir el incentivo económico en el presente mes por estar suspendido de manera temporal.

2.3.5.5. Prestaciones directas a los hogares beneficiados

Constituido por los servicios de entrega del incentivo monetario a los hogares beneficiados así como de la articulación con la oferta para facilitar el acceso a los establecimientos de salud, nutrición, educación e identidad.

2.3.6. Control de condicionalidades

2.3.6.1. Condicionalidad

Compromiso contraído por el representante del hogar beneficiado con el PNADP, mediante el cual los miembros objetivo, deberán asistir a los servicios básicos de nutrición, salud y educación

2.3.6.2. Formato FE-002

Formato de certificación de incumplimiento de asistencia a primaria, en el cual se registra mensualmente el número de inasistencias a la institución educativa, por parte del miembro objetivo con una medición trimestral.

2.3.6.3. Formato FM-001

Formato de registro de incumplimiento de asistencia a la entrega del complemento nutricional, en el cual se registra mensualmente el número de inasistencias al recojo del complemento nutricional, con una medición trimestral.

2.3.6.4. Formato FS-002

Ficha familiar, en el cual se registra el cumplimiento mensual de los controles programados y las asistencias a los establecimientos de salud, por parte de los miembros objetivos.

2.3.6.5.Formato FS-003

Documento de registro de incumplimiento de asistencia a los servicios de salud, en el cual se registra el mes o meses del trimestre en que algún integrante del hogar objetivo no asiste a alguno de los controles programados

2.3.6.6.Hogar suspendido

Corresponde a un hogar beneficiado que se encuentra privado del incentivo monetario en razón del incumplimiento de los compromisos asumidos con el PNADP; esta acción de carácter temporal.

2.3.6.7.Miembros de hogar sujetos al sector nutrición

Integrantes del Hogar cuya edad está entre los 06 meses a 3 años, inscritos en un establecimiento de salud,

2.3.6.8.Miembros del hogar sujetos al sector salud

Integrantes del Hogar cuya edad está entre los 0 a 14 años de edad, inscritos en un establecimiento de salud, comprende también a la madre gestante.

2.3.6.9.Miembros del hogar sujetos al sector educación

Integrantes del Hogar cuya edad está entre 6 a 14 años de edad y pertenece a un grado de primaria.

2.3.7. En relación a la determinación de los representantes de hogares beneficiados sujetos a pago de los fondos condicionados

2.3.7.1. Incentivo económico-monetario

Monto determinado en Cien Nuevos Soles, que constituye el fondo condicionado transferido mensualmente a los hogares beneficiados.

2.3.7.2. Apertura de cuentas

Proceso mediante el cual el PNADP gestiona ante el Banco de la Nación la apertura de las cuentas de ahorro a nombre de cada representante de hogar afiliado.

2.3.7.3. Abono

Operación bancaria efectuada por el Banco de la Nación en atención a la transferencia de fondos realizados por el PNADP. Esta operación se desdobra en dos modalidades: Abonos por agencia bancaria y por transportadora.

2.3.7.4. Oficina bancaria

Oficina descentralizada del Banco de la Nación.

2.3.7.5. Transportadora

Empresa transportadora de valores, contratada por el Banco de la Nación para atender a los beneficiados en las zonas donde no hay oficina bancaria.

2.3.8. En relación a la gestión de los procesos de la GOAL

2.3.8.1. Procesos

Los 3 grandes procesos del PNADP son: Afiliaciones, Control de cumplimiento de compromisos y Pago.

2.3.8.2. Clientes

Clientes internos: las gerencias, los promotores, las coordinaciones regionales. Clientes externos: los hogares beneficiados, los órganos de control externo, las autoridades locales, regionales, el consejo directivo.

2.3.8.3. Productos

Listado de padrón de abonos, informes de afiliación, transferencias a las cuentas de los titulares beneficiarios.

2.3.8.4. Proveedores

Los proveedores del PNADP son: RENIEC (Cotejo de registro masivo de DNI, entrega de las imágenes de las fotografías de los titulares beneficiarios); INEI (Barrido censal, Base de datos); Banco de la Nación (Servicio de depósitos de los titulares beneficiarios y Servicio de pagaduría), las entidades de los Sectores MINSA, MIMDES, MINEDU que se relacionan a la oferta de los servicios básicos a entregar a los hogares beneficiados.

CAPITULO III

3. DESARROLLO DE SOLUCIÓN

El presente capítulo desarrolla la solución propuesta. El desarrollo contempla el diseño del plan del proyecto del Datawarehouse así como la arquitectura propuesta, de acuerdo a la metodología planteada por Ramon Barquin. En las primeras secciones se hace un análisis de los factores que intervendrán en la solución y se recopilan los requerimientos de negocio. En las secciones 3.5, 3.6 y 3.7 se explica a detalle la arquitectura, los procesos de carga y extracción, así como las consideraciones en lo referente a la calidad de datos. Finalmente, se incluye un plan de implementación y testeo para el *Datawarehouse*.

3.1 Desarrollo del Plan

3.1.1. Misión

Diseñar e Implementar una solución de Inteligencia de negocios (BI) para el PNADP. Incluirá el diseño de un *Datawarehouse* y un tablero de control.

3.1.2. Análisis de riesgos

La figura 32 muestra los riesgos y amenazas que se han identificado frente a los objetivos específicos de la solución propuesta.

Objetivos específicos de la solución propuesta	Riesgos/Amenazas/Vulnerabilidad
Lograr una percepción integral de las necesidades de análisis de información del PNADP.	<p>Riesgo: Los requerimientos obtenidos no son integrales.</p> <p>Amenaza: Percepción sesgada de los analistas funcionales.</p> <p>Vulnerabilidad: Cada analista funcional tiene un alto conocimiento sólo de sus operaciones.</p>
Proporcionar las herramientas adecuadas para permitir a los gestores el análisis detallado de la información de los hogares beneficiados del PNADP.	<p>Riesgo: Las herramientas que se diseñan no facilitan el nivel de análisis deseable.</p> <p>Amenaza: Diseño inadecuado de la solución.</p> <p>Vulnerabilidad: Las premuras en los tiempos no permiten la participación de los usuarios en las fases tempranas de diseño.</p>
Implementar gráficamente los principales indicadores de negocio del PNADP.	<p>Riesgo: Indicadores de alto nivel no cumplen las expectativas planteadas.</p> <p>Amenaza: Los indicadores implementados no son los adecuados.</p> <p>Vulnerabilidad: Poca interacción con el personal de mas alto nivel.</p>
Implementar los procesos de administración adecuados para la sostenibilidad de la solución a implementar.	<p>Riesgo: No se definen los procesos de administración adecuados.</p> <p>Amenaza: La solución implementada no tiene sostenibilidad.</p> <p>Vulnerabilidad: Procesos poco definidos.</p>

Figura 32: Riesgos y Amenazas frente a los objetivos de un Datawarehouse

La figura 33 muestra el análisis de los riesgos identificados y una propuesta de acción frente a los riesgos.

Riesgo	Probabilidad	Impacto	Acción
Los requerimientos obtenidos no son integrales.	Baja	Alto	Realizar reuniones con analistas funcionales de todas las áreas.
Las herramientas que se diseñan no facilitan el nivel de análisis deseable.	Baja	Medio	Elaborar prototipos e involucrar a los usuarios claves.
Indicadores de alto nivel no cumplen las expectativas planteadas.	Media	Alto	Análisis detallado de los planes operativos.
No se definen los procesos de administración adecuados.	Alta	Alto	Definición de roles de administración del proyecto. Aprovechar el proyecto de implementación del Plan de seguridad del PNADP.

Figura 33: Análisis de Riesgos

3.1.3. Definición de roles para la administración del proyecto

Para la exitosa implementación del Datawarehouse, se están considerando los siguientes roles:

- Rol1: Gestor de proyecto del Datawarehouse
- Rol2: Arquitecto del Datawarehouse
- Rol3: Administrador de la base de datos
- Rol4: Especialista en migración de datos
- Rol5: Especialista en la transformación y filtrado de datos

- Rol6: Líder de desarrollo
- Rol7: Especialista en calidad y pruebas
- Rol8: Usuarios clave
- Rol9: Especialista en herramientas
- Rol10: Entrenador
- Rol11: Funcionalidad de Asistencia al usuario
- Rol12: Especialista en provisión de datos

Se propone una estrategia de asignación de roles, la cual se muestra en la figura 34. Se puede notar que los roles de Entrenador y Funcionalidad de asistencia al usuario se han derivado al área de operaciones. Esto es debido a que se busca el empoderamiento de los usuarios en el *Datawarehouse*.

Gerencia	Rol
Tecnología de información	Rol1, Rol2, Rol3, Rol4, Rol5, Rol6, Rol7, Rol9, Rol12
GOAL	Rol8, Rol10, Rol11
Alta dirección	Rol8

Figura 34: Estrategia de asignación de Roles

3.1.4. Equipo del Proyecto

Se han identificado a 8 miembros del equipo, cada uno tiene más de un rol. En la en la figura 35 especifican los roles de cada uno.

Miembros del Equipo del Proyecto	Roles
Gerente de la gerencia de Tecnología de información.	Rol1
Jefe de proyecto	Rol2, Rol6
Analista de Sistemas I	Rol2, Rol5, Rol7, Rol12
Analista de Sistemas II	Rol9, Rol10
Administrador de Base de Datos	Rol3, Rol4
Analista de información del proceso de Afiliaciones	Rol8, Rol11
Analista de información del proceso de Condicionalidades	Rol8, Rol11
Analista de información del proceso de Pagos	Rol8, Rol11

Figura 35: Equipo del proyecto

3.1.5. Recursos disponibles para la implementación del proyecto

3.1.5.1 Documentación

- Plan Operativo 2007.
- Diccionario de datos BD producción
- Hojas de operaciones GOAL.

3.1.5.2 Tecnología: Software

- BD ORACLE 10g release 2.
- ORACLE Discoverer 10g.
- ORACLE Portal 10g.
- Herramienta de Modelamiento de datos.
- 1 Servidor de BD Stand By.
- 1 Servidor de Test.
- 1 Servidor de Aplicaciones
- Conexión de red fiable.

3.1.5.3 Tecnología: Hardware

Configuración Servidor Stand By:

- Discos Físicos: 2 discos de 140GB
- Raid: 2xR1
- Espacio real disponible: 140GB
- Procesador: 2x3.4 Xeon
- Memoria RAM: 8GB
- Memoria SWAP: 8GB
- Sistema Operativo: Red Hat Enterprise Linux Release 4
update 3

3.2 Identificación de Requerimientos

La identificación de requerimientos ha sido el producto de entrevistas con los usuarios clave del equipo del proyecto, los cuales también han colaborado con la entrega de documentación histórica de indicadores de gestión útiles para medir los resultados de los procesos.

Uno de los requerimientos es que los indicadores de gestión se deben presentar de manera mensual ya que la mayoría de los procesos se ejecutan mensualmente; sin embargo hay algunos procesos que se ejecutan diariamente.

Los resultados generales de los indicadores de gestión también se deben mostrar agrupados por censo, distrito y departamento.

3.2.1. Indicadores de Gestión para el proceso de Afiliaciones

3.2.1.1 Subproceso de Carga de información de Censos

- REQACE01: Número de hogares censados
- REQACE02: Número de hogares censados pobres
- REQACE03: Porcentaje de hogares pobres con respecto al total de censados
- REQACE04: Número de hogares censados no pobres
- REQACE05: Porcentaje de hogares no pobres con respecto al total de censados
- REQACE06: Número de miembros de hogares censados

- REQACE07: Número de miembros de hogares censados pobres
- REQACE08: Porcentaje de miembros de hogares pobres con respecto al total de censados
- REQACE09: Número de miembros de hogares censados no pobres
- REQACE10: Porcentaje de miembros de hogares no pobres con respecto al total de censados
- REQACE11: Número de niños de 0 a 5 años de hogares censados
- REQACE12: Número de niños de 0 a 5 años de hogares censados pobres
- REQACE13: Porcentaje de niños de 0 a 5 años de hogares pobres con respecto al total de censados
- REQACE14: Número de niños de 0 a 5 años de hogares censados no pobres
- REQACE15: Porcentaje de niños de 0 a 5 años de hogares no pobres con respecto al total de censados
- REQACE16: Número de niños de 6 a 14 años de hogares censados
- REQACE17: Número de niños de 6 a 14 años de hogares censados pobres
- REQACE18: Porcentaje de niños de 6 a 14 años de hogares censados pobres con respecto al total de censados

- REQACE19: Número de niños de 6 a 14 años de hogares censados no pobres
- REQACE20: Porcentaje de niños de 6 a 14 años de hogares censados no pobres con respecto al total de censados
- REQACE21: Número de gestantes de hogares censados
- REQACE22: Número de gestantes de hogares censados pobres
- REQACE23: Porcentaje de gestantes de hogares pobres con respecto al total de censados
- REQACE24: Número de gestantes de hogares censados no pobres
- REQACE25: Porcentaje de gestantes de hogares no pobres con respecto al total de censados
- REQACE26: Porcentaje de viviendas por tipo de material de pared
- REQACE27: Porcentaje de viviendas por tipo de material de techo
- REQACE28: Porcentaje de viviendas por tipo de material de pisos
- REQACE29: Porcentaje de viviendas por tipo de alumbrado
- REQACE30: Porcentaje de viviendas por tipo de abastecimiento de agua.

3.2.1.2 Subproceso de Validación por taller

- REQAVT01: Número de hogares pendientes de validar por taller
- REQAVT02: Porcentaje de hogares pendientes de validar con respecto al total de hogares
- REQAVT03: Número de hogares validados por taller
- REQAVC04: Porcentaje de hogares validados con respecto al total de hogares
- REQAVC05: Número de hogares no validados por taller
- REQAVC06: Porcentaje de hogares no validados con respecto al total de hogares
- REQAVC07: Número de hogares validados que no tienen DNI

3.2.1.3 Subproceso de Envío a Cotejo RENIEC

- REQACR01: Número de hogares enviados a Cotejo de RENIEC
- REQACR02: Número de hogares enviados a Cotejo de RENIEC sin restricción
- REQACR03: Porcentaje de hogares sin restricción con respecto al total enviados a cotejo de RENIEC
- REQACR04: Número de hogares enviados a Cotejo de RENIEC con restricción
- REQACR05: Porcentaje de hogares con restricción con respecto al total enviados a cotejo de RENIEC

- REQACR06: Número de hogares enviados a Cotejo de RENIEC con restricción validados por GOAL
- REQACR07: Porcentaje de hogares con restricción validados por GOAL con respecto al total enviados a cotejo de RENIEC

3.2.1.4 Subproceso de Firma de Convenio

- REQAFC01: Número de hogares con firma de convenio
- REQAFC02: Porcentaje de hogares con firma de convenio con respecto al total
- REQAFC03: Número de hogares sin firma de convenio
- REQAFC04: Porcentaje de hogares sin firma de convenio con respecto al total

3.2.1.5 Subproceso de Suspensiones Temporales

- REQAST01: Número de hogares con suspensión temporal
- REQAST02: Porcentaje de hogares con suspensión temporal por motivo de suspensión

3.2.1.6 Subproceso de Exclusiones

- REQAEX01: Número de hogares excluidos del PNADP
- REQAEX02: Porcentaje de hogares excluidos por motivo
- REQAEX03: Porcentaje de hogares excluidos con respecto al total de censados
- REQAEX04: Número de hogares excluidos con abonos.

3.2.1.7 Subproceso de Cambio de Titular

- REQACT01: Número de hogares que han hecho cambio de titular
- REQACT02: Porcentaje de hogares con cambio de titular por motivo

3.2.1.8 Información general del proceso de Afiliaciones

- REQAIG01: Número hogares activos
- REQAIG02: Porcentaje de hogares activos con respecto al total de hogares censados
- REQAIG03: Número de hogares afiliados
- REQAIG04: Porcentaje de hogares afiliados con respecto al total de hogares activos
- REQAIG05: Número de hogares no afiliados
- REQAIG06: Porcentaje de hogares no afiliados con respecto al total de hogares activos
- REQAIG07: Porcentaje de hogares no afiliados por motivo
- REQAIG08: Número de Niños que no tienen fecha de nacimiento
- REQAIG09: Porcentaje de Niños que no tienen fecha de nacimiento con respecto al total
- REQAIG10: Número de Niños que no tienen partida de nacimiento
- REQAIG11: Porcentaje de Niños que no tienen partida de nacimiento con respecto al total

- REQAIG12: Número de miembros del hogar mayores de 15 años que son analfabetas.
- REQAIG13: Porcentaje de miembros de hogar mayores de 15 años que son analfabetas con respecto al total

3.2.2. Indicadores de Gestión para el proceso de condicionalidades

3.2.2.1 Subproceso de Incorporación a los Sectores

- REQCIS01: Número de hogares no incorporados en Salud
- REQCIS02: Porcentaje de hogares no incorporados en Salud con respecto al total
- REQCIS03: Número de niños de 0 a 14 años y/o gestantes no incorporados en Salud
- REQCIS04: Porcentaje de niños de 0 a 14 años y/o gestantes no incorporados en Salud con respecto al total
- REQCIS05: Número de hogares con niños en edad Escolar
- REQCIS06: Porcentaje de hogares con niños en edad escolar con respecto al total
- REQCIS07: Número de Niños en edad Escolar
- REQCIS08: Porcentaje de niños en edad escolar con respecto al total
- REQCIS09: Hogares no incorporados en Educación
- REQCIS10: Porcentaje de hogares no incorporados en Educación con respecto al total de hogares con niños en edad escolar

- REQCIS11: Niños no incorporados en Educación
- REQCIS12: Porcentaje de niños no incorporados en Educación con respecto al total de niños en edad escolar
- REQCIS13: Número de hogares con niños de en edad de recibir papilla
- REQCIS14: Porcentaje de hogares con niños en edad de recibir papilla con respecto al total
- REQCIS15: Número de niños en edad de recibir papilla
- REQCIS16: Porcentaje de niños en edad de recibir papilla con respecto al total
- REQCIS17: Porcentaje de niños no incorporados en Salud con respecto al total de niños en edad de recibir papilla

3.2.2.2 Subproceso de generación de formatos

- REQCGF01: Número de hogares que se esperaba evaluar en Salud
- REQCGF02: Número de hogares que se esperaba evaluar en Educación
- REQCGF03: Número de hogares que se esperaba evaluar en Nutrición
- REQCGF04: Número de niños y/o gestantes que se esperaba evaluar en Salud
- REQCGF05: Número de niños que se esperaba evaluar en Educación

- REQCGF06: Número de niños que se esperaba evaluar en Nutrición
- REQCGF07: Número de hogares no evaluados en Salud
- REQCGF08: Porcentaje de hogares no evaluados en Salud con respecto al total.
- REQCGF09: Número de niños y/o gestantes no evaluados en Salud
- REQCGF10: Porcentaje de niños y/o gestantes no evaluados en Salud con respecto al total.
- REQCGF11: Número de hogares no evaluados en Educación
- REQCGF12: Porcentaje de hogares no evaluados en Educación con respecto al total.
- REQCGF13: Número de niños no evaluados en Educación
- REQCGF14: Porcentaje de niños no evaluados en Educación con respecto al total.
- REQCGF15: Número de hogares no evaluados en Nutrición
- REQCGF16: Porcentaje de hogares no evaluados en Nutrición con respecto al total.
- REQCGF17: Número de niños no evaluados en Nutrición
- REQCGF18: Porcentaje de niños no evaluados en Nutrición con respecto al total.
- REQCGF19: Número de Formatos generados en Educación.

- REQCGF20: Número de Formatos generados en Salud.
- REQCGF21: Número de Formatos generados en Nutrición.

3.2.2.3 Subproceso de evaluación de cumplimiento de compromisos

- REQCEC01: Número de hogares evaluados en Salud
- REQCEC02: Número de hogares que no cumplieron en Salud
- REQCEC03: Porcentaje de hogares que no cumplieron en Salud con respecto al total de evaluados.
- REQCEC04: Número de hogares evaluados en Educación
- REQCEC05: Número de hogares que no cumplieron en Educación
- REQCEC06: Porcentaje de hogares que no cumplieron en Educación con respecto al total de evaluados.
- REQCEC07: Número de hogares evaluados en Nutrición
- REQCEC08: Número de hogares que no cumplieron en Nutrición
- REQCEC09: Porcentaje de hogares que no cumplieron en Nutrición con respecto al total de evaluados.
- REQCEC10: Número de niños evaluados en Educación
- REQCEC11: Número de niños que no cumplieron en Educación
- REQCEC12: Porcentaje de niños que no cumplieron en Educación con respecto al total de evaluados.
- REQCEC13: Número de hogares evaluados en Nutrición

- REQCEC14: Número de niños que no cumplieron en Nutrición
- REQCEC15: Porcentaje de niños que no cumplieron en Nutrición con respecto al total de evaluados.
- REQCEC16: Número de hogares suspendidos por incumplimiento de compromisos
- REQCEC17: Porcentaje de hogares suspendidos con respecto al total de evaluados

3.2.2.4 Información General del proceso de Condicionalidades

- REQCIG01: Número de hogares evaluados que no tienen ningún incumplimiento de compromisos por cantidad de evaluaciones
- REQCIG02: Número de hogares evaluados con mayor reincidencia de incumplimiento
- REQCIG03: Número de centros educativas con mayor incumplimiento de compromisos
- REQCIG04: Número de Centros de Salud con mayor incumplimiento de compromisos
- REQCIG05: Número de centros educativas que no tienen ningún incumplimiento de compromisos
- REQCIG06: Número de centros de salud que no tienen ningún incumplimiento de compromisos

3.2.2.5 Información General de las instituciones

- REQCI01: Número de centros educativos registradas.

- REQCII02: Número de centros educativos dentro de la cobertura del PNADP.
- REQCII03: Número centros educativos con al menos un miembro objetivo incorporado.
- REQCII04: Número centros educativos fuera de la cobertura del PNADP y con al menos un miembro objetivo incorporado.
- REQCII05: Número establecimientos de salud registradas.
- REQCII06: Número establecimientos de salud dentro de la cobertura del PNADP.
- REQCII07: Número establecimientos de salud con al menos un miembro objetivo incorporado.
- REQCII08: Número establecimientos de salud fuera de la cobertura del PNADP y con al menos un miembro objetivo incorporado.

3.2.3. Indicadores de Gestión para el proceso de Pagos

3.2.3.1 Subproceso de Apertura de cuentas

- REQPA01: Número de hogares que no tienen cuenta
- REQPA02: Número de hogares que se les solicita apertura de cuentas
- REQPA03: Porcentaje de hogares que se les solicita apertura de cuentas con respecto al total que no tienen cuentas

- REQPA04: Número de hogares que no se les abrió una cuenta.
- REQPA05: Porcentaje de hogares que no se les abrió cuenta con respecto al total que se le solicitó apertura de cuenta.

3.2.3.2 Subproceso de Cierre de Padrón de Abonos

- REQPA01: Número de hogares en el padrón Activo
- REQPA02: Porcentaje de hogares en el padrón Activo con respecto al total de hogares validados por Taller.
- REQPA03: Número de hogares en el padrón Pasivo
- REQPA04: Porcentaje de hogares en el padrón Pasivo con respecto al total de hogares validados por Taller.
- REQPA05: Número de hogares que no están en el Padrón de Abonos por no estar afiliados
- REQPA06: Porcentaje de hogares no afiliados con respecto al total de hogares validados por Taller.
- REQPA07: Número de hogares que están en el Pasivo por incumplimiento de compromisos
- REQPA08: Porcentaje de hogares que están en el Pasivo por incumplimiento de compromisos con respecto al total de hogares validados por Taller.
- REQPA09: Número de hogares que están en el Pasivo por no estar incorporados en Salud

- REQPPA10: Porcentaje de hogares que están en el Pasivo por no estar incorporados en Salud con respecto al total de hogares validados por Taller.
- REQPPA11 Número de hogares que están en el Pasivo por no estar incorporados en Educación
- REQPPA12: Porcentaje de hogares que están en el Pasivo por no estar incorporados en Educación con respecto al total de hogares validados por Taller.
- REQPPA13: Número de hogares que entraron al padrón de abonos por primera vez
- REQPPA14: Número de hogares que entraron al padrón de abonos el mes actual con respecto al padrón de abonos del mes anterior.
- REQPPA15: Número de hogares que salieron del padrón de abonos el mes actual con respecto al padrón de abonos del mes anterior.

3.2.3.3 Subproceso de Puntos de Pago

- REQPPP01: Número de hogares en el padrón activo (hogares beneficiados) por modalidad de pago
- REQPPP02: Número de hogares beneficiados con modalidad Transportadora
- REQPPP03: Porcentaje de hogares con modalidad por transportadora con respecto al total de hogares beneficiados

- REQPPP04: Número de hogares beneficiados con modalidad agencia bancaria
- REQPPP05: Porcentaje de hogares con modalidad por agencia bancaria con respecto al total de hogares beneficiados
- REQPPP06: Número de Puntos de pago por Distrito

3.2.3.4 Información General del proceso de Pagos

- REQPIG01: Número de hogares beneficiados por primera vez.
- REQPIG02: Número de hogares con más de 12 abonos ininterrumpidos
- REQPIG03: Número de hogares que llegan al máximo de cantidad de abonos con respecto a hogares del mismo censo
- REQPIG04: Número de hogares que están debajo del promedio de cantidad de abonos con respecto a hogares del mismo censo
- REQPIG05: Número de hogares beneficiados y que no se les ha evaluado los compromisos.
- REQPIG06: Porcentaje de hogares beneficiados y que no se les ha evaluado los compromisos con respecto al total de hogares beneficiados.
- REQPIG07: Número de niños de 0 a 5 años que pertenecen a hogares beneficiados.

- REQPIG08: Porcentaje de niños de 0 a 5 años que pertenecen a hogares beneficiados con respecto al total de niños y/o gestantes de hogares beneficiados.
- REQPIG09: Número de niños de 6 a 14 años que pertenecen a hogares beneficiados.
- REQPIG10: Porcentaje de niños de 6 a 14 años que pertenecen a hogares beneficiados con respecto al total de niños y/o gestantes de hogares beneficiados.

3.3 Modelamiento de Datos

Se han definido los modelos estrella, en base al análisis de los hechos (Fact tables) y las dimensiones requeridas. El detalle de los modelos se muestra a continuación:

3.3.1. Modelo Estrella: Identificación de población censada

Este modelo estrella describe las principales características de la población censada (el hecho), analizada en base al ubigeo, umbral de pobreza, la etapa y el tiempo. La dimensión población se incluye para obtener el detalle de la población censada. Con este modelo se pueden atender los siguientes requerimientos: REQACE01, REQACE06, REQACE11, REQACE16, REQACE21.

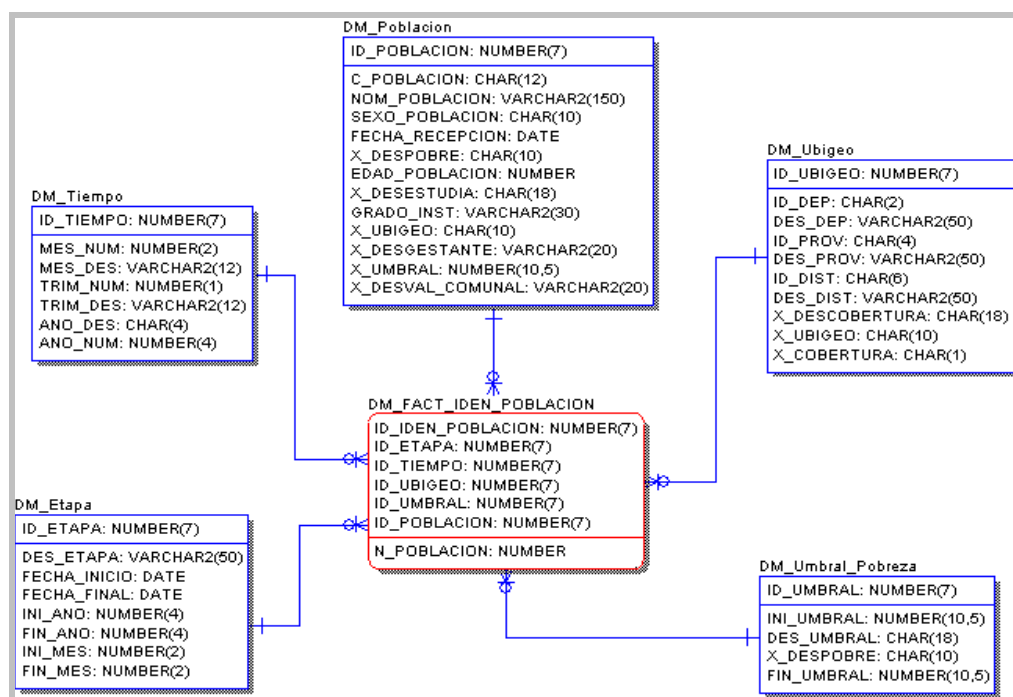


Figura 36: Modelo Estrella – Identificación de Población censada

3.3.2. Modelo Estrella: Identificación población censada pobre

Este modelo estrella describe las principales características de la población censada pobre (el hecho), analizada en base al ubigeo, umbral de pobreza, la etapa y el tiempo. La dimensión población no pobre se incluye para obtener el detalle de la población censada pobre. Con este modelo se pueden atender los siguientes requerimientos: REQACE02, REQACE03, REQACE07, REQACE08, REQACE12, REQACE13, REQACE17, REQACE18, REQACE22, REQACE23.

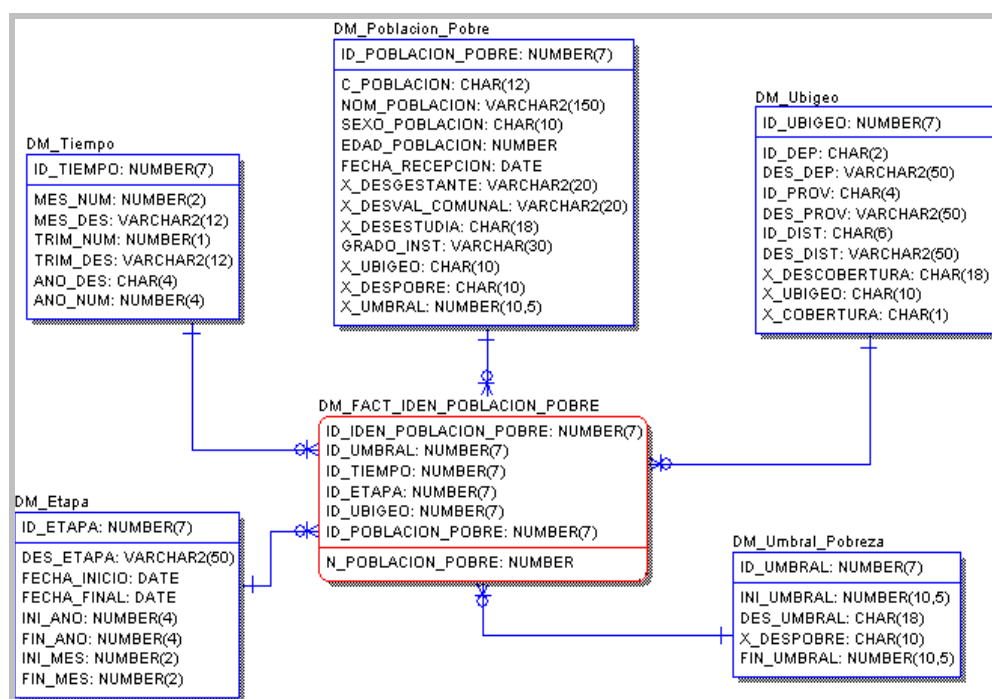


Figura 37: Modelo Estrella – Identificación de Población censada pobre

3.3.3. Modelo Estrella: Identificación población censada no pobre

Este modelo estrella describe las principales características de la población censada no pobre (el hecho), analizada en base al ubigeo, umbral de pobreza, la etapa y el tiempo. La dimensión población no pobre se incluye para obtener el detalle de la población censada no pobre. Con este modelo se pueden atender los siguientes requerimientos: REQACE04, REQACE05, REQACE09, REQACE10, REQACE14, REQACE15, REQACE19, REQACE20, REQACE24, REQACE25

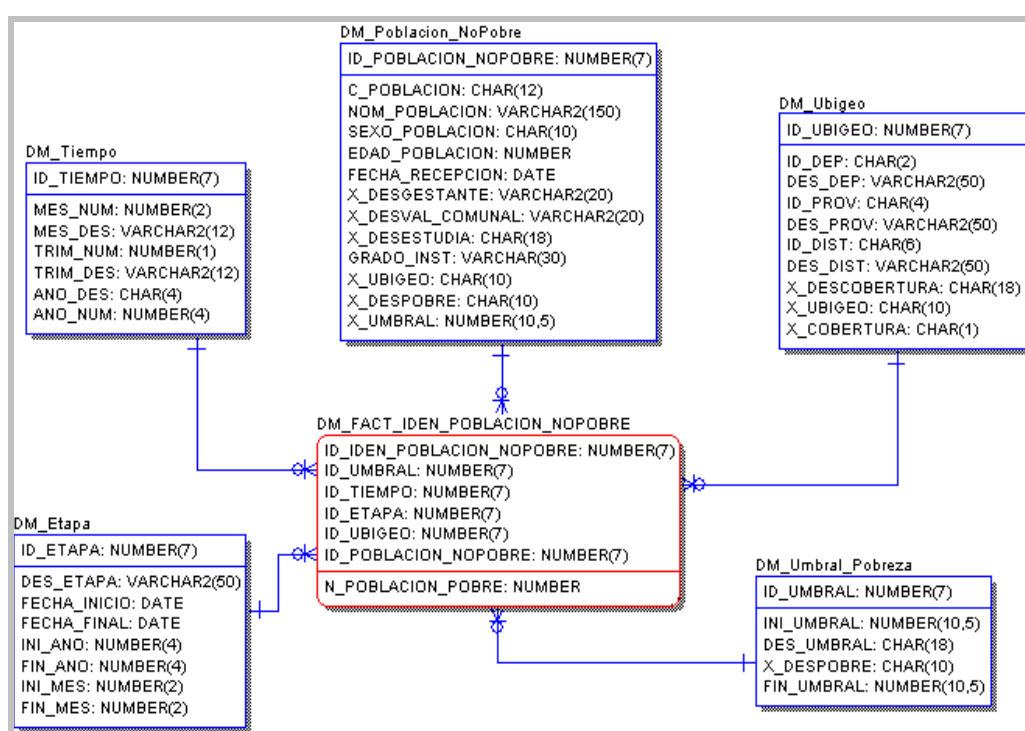


Figura 38: Modelo Estrella – Identificación de Población censada pobre

3.3.4. Modelo Estrella: Identificación de vivienda

Este modelo estrella describe las principales características de las viviendas censadas (el hecho) analizada en base al ubigeo, umbral de pobreza, la etapa y el tiempo. La dimensión vivienda se incluye para obtener el detalle de las viviendas censadas. Con este modelo se pueden atender los siguientes requerimientos: REQACE26, REQACE27, REQACE28, REQACE29, REQACE30

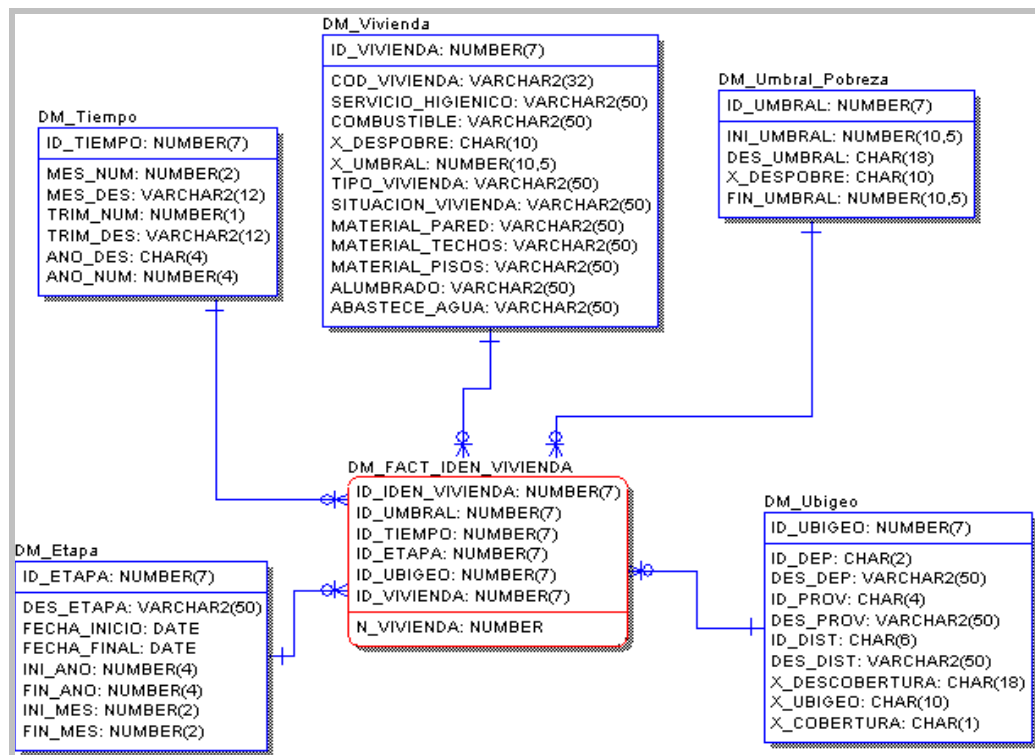


Figura 39: Modelo Estrella – Identificación de Vivienda

3.3.5. Modelo Estrella: Identificación del hogar afiliado

Este modelo estrella describe las principales características de los hogares afiliados y no afiliados para verificar su estado de afiliación (el hecho) analizada en base al ubigeo, tipo de suspensión, tipo de padrón (activo, pasivo, etc.) y el tiempo. La dimensión titular se incluye para obtener los datos del titular. Con este modelo se pueden atender los siguientes requerimientos: REQAVT01, REQAVT02, REQAVT03, REQAVT04, REQAVT05, REQAVT06, REQAVT07, REQAFC01, REQAFC02, REQAFC03, REQAFC04, REQAST01, REQAST02, REQAEX01, REQAEX02, REQAEX03, REQAEX04, REQACT01, REQACT02, REQAIG01, REQAIG03, REQAIG04, REQAIG05, REQAIG06, REQAIG07.

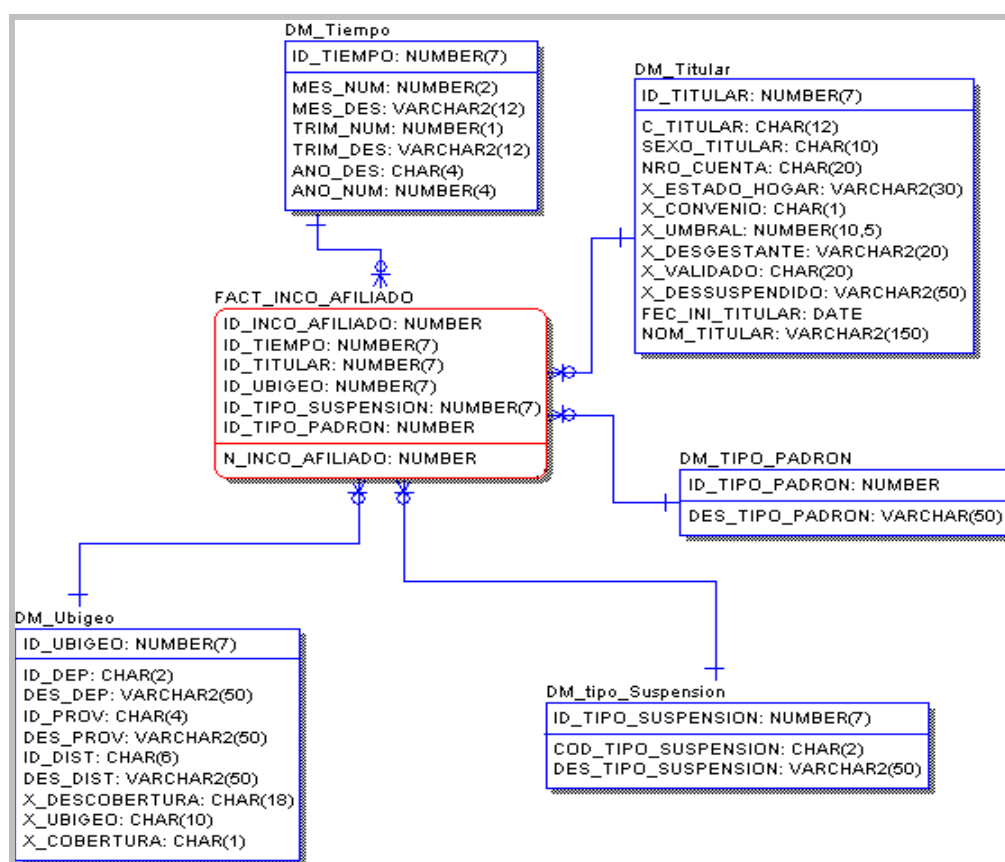


Figura 40: Modelo Estrella – Identificación del hogar afiliado

3.3.6. Modelo Estrella: Identificación de los miembros objetivos

Este modelo estrella describe las principales características de los miembros objetivos (beneficiarios) que pertenecen a hogares afiliados y no afiliados (el hecho) analizada en base al ubigeo, el tiempo, etapa, titular del hogar y rango de edad. La dimensión titular se incluye para obtener los datos del titular. Con este modelo se pueden atender los siguientes requerimientos:

REQAIG08, REQAIG09, REQAIG10, REQAIG11

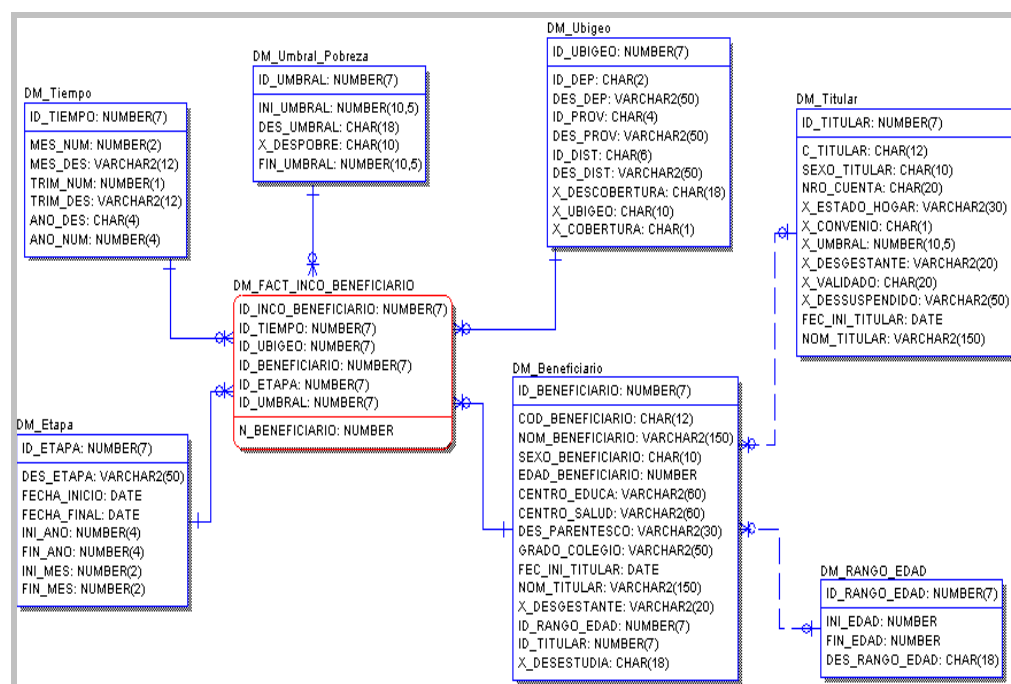


Figura 41: Modelo Estrella – Identificación de los miembros objetivos

3.3.7. Modelo Estrella: Identificación de Centro Educativos

Este modelo estrella describe las principales características de los centros educativos registrados en el PNADP (el hecho) analizada en base al ubigeo. La dimensión centro educa se incluye para obtener el detalle de los centros educativos. Con este modelo se pueden atender los siguientes requerimientos: REQCII01, REQCII02, REQCII03, REQCII04.

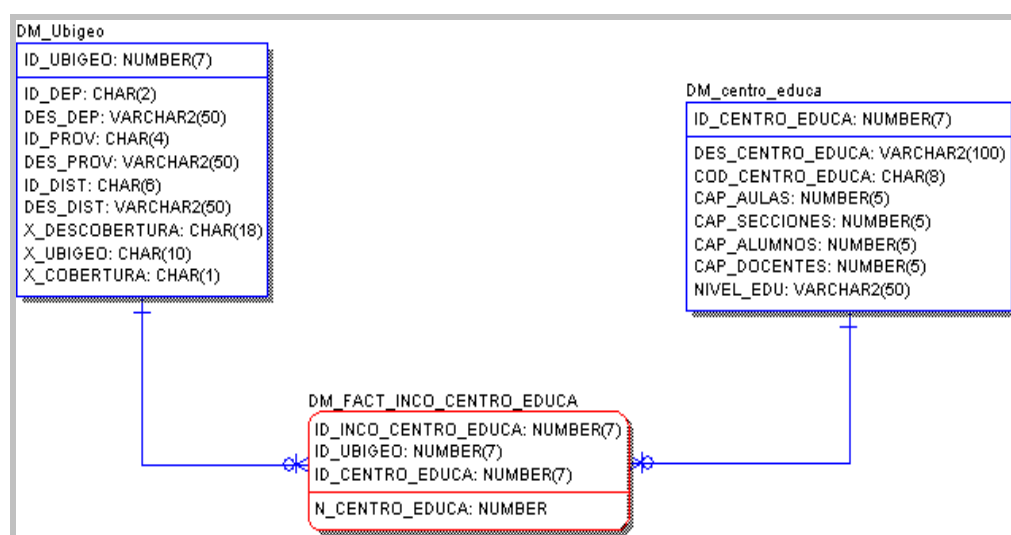


Figura 42: Modelo Estrella – Identificación de Centros Educativos

3.3.8. Modelo Estrella: Identificación de Establecimientos de Salud

Este modelo estrella describe las principales características de los establecimientos de salud registrados en el PNADP (el hecho) analizada en base al ubigeo. La dimensión centro salud se incluye para obtener el detalle de los establecimientos de salud. Con este modelo se pueden atender los siguientes requerimientos: REQCII05, REQCII06, REQCII07, REQCII08.

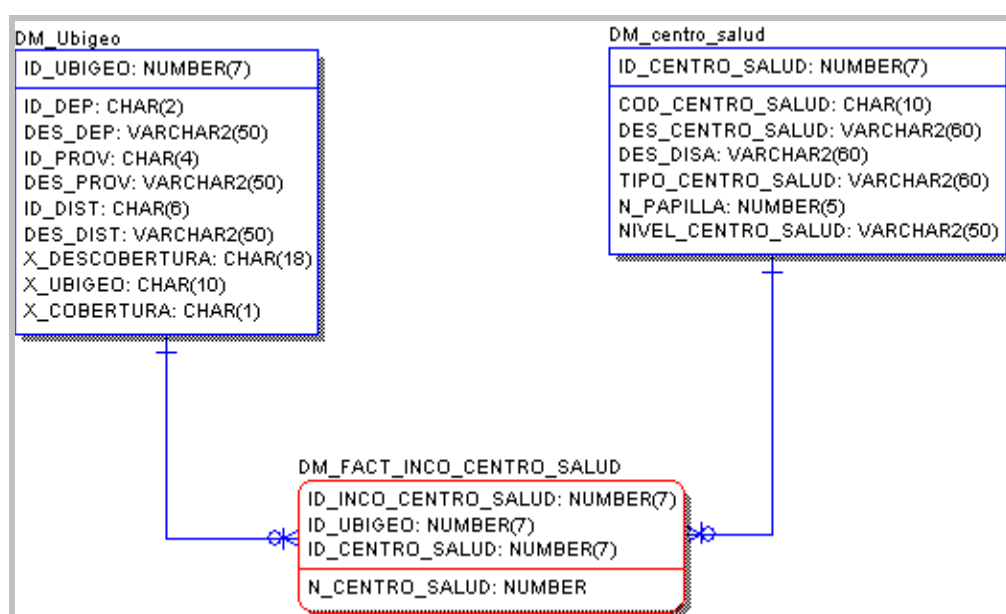


Figura 43: Modelo Estrella – Identificación de Establecimientos de salud

3.3.9. Modelo Estrella: Identificación de Hogares incorporados

Este modelo estrella describe las principales características de hogares incorporados al PNADP (el hecho) analizada en base al ubigeo, etapa, umbral de pobreza y el tiempo. La dimensión hogar se incluye para obtener el detalle de los hogares incorporados al PNADP. Con este modelo se pueden atender los siguientes requerimientos: REQCIS01, REQCIS02, REQCIS03, REQCIS04, REQCIS09, REQCIS10, REQCIS11, REQCIS12, REQCIS17.

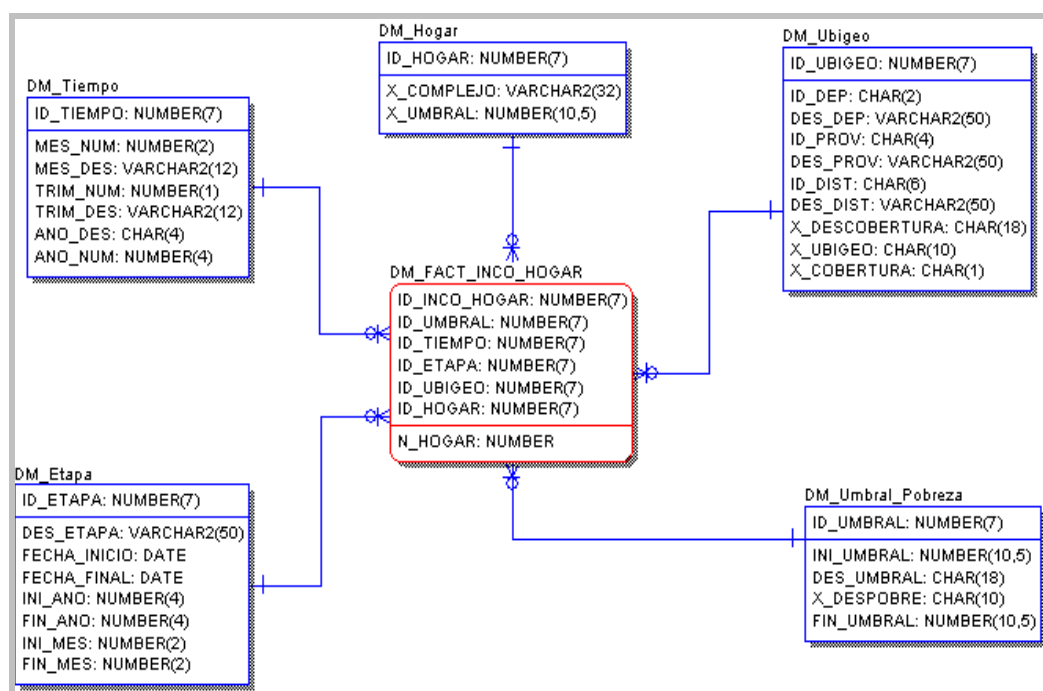


Figura 44: Modelo Estrella – Identificación de Hogares incorporados

3.3.10. Modelo Estrella: Identificación de Promotores

Este modelo estrella describe las principales características de los promotores sociales del PNADP (el hecho) analizada en base al ubigeo y el tipo de promotor. La dimensión promotor se incluye para obtener el detalle de los promotores sociales del PNADP.

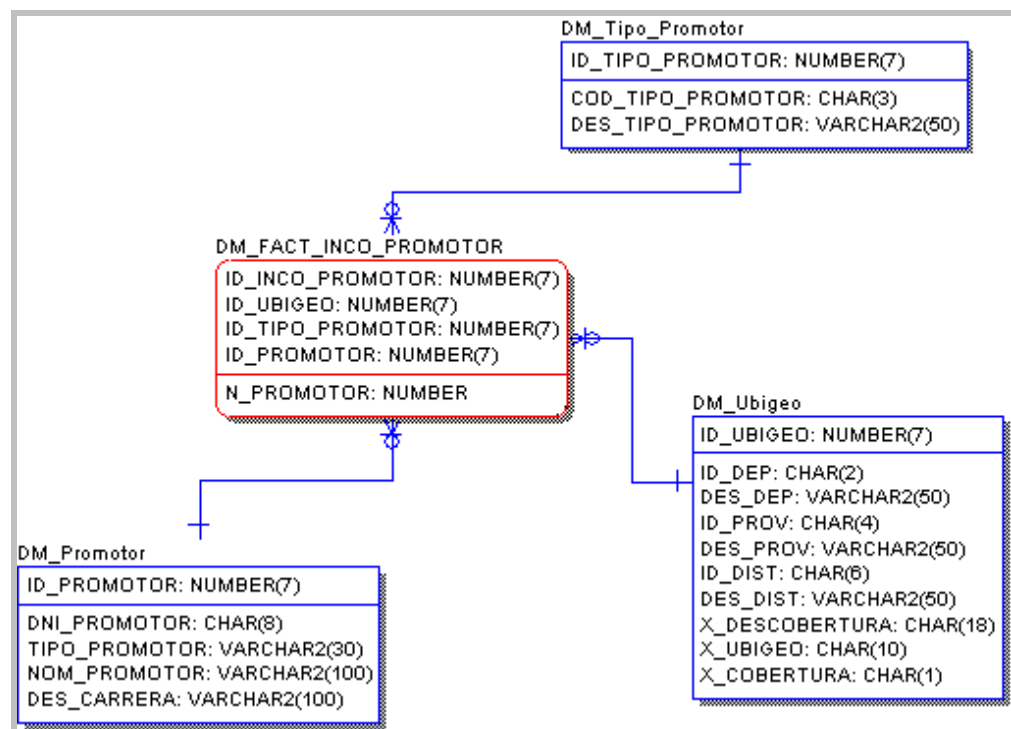


Figura 45: Modelo Estrella – Identificación de Promotores

3.3.11. Modelo Estrella: Identificación de Hogares validados

Este modelo estrella describe las principales características de los titulares validados en asamblea (el hecho) analizada en base al ubigeo y al tiempo. La dimensión titular se incluye para obtener el detalle de los titulares validados en asamblea. REQCIS05, REQCIS06, REQCIS07, REQCIS08, REQCIS13, REQCIS14, REQCIS15, REQCIS16.

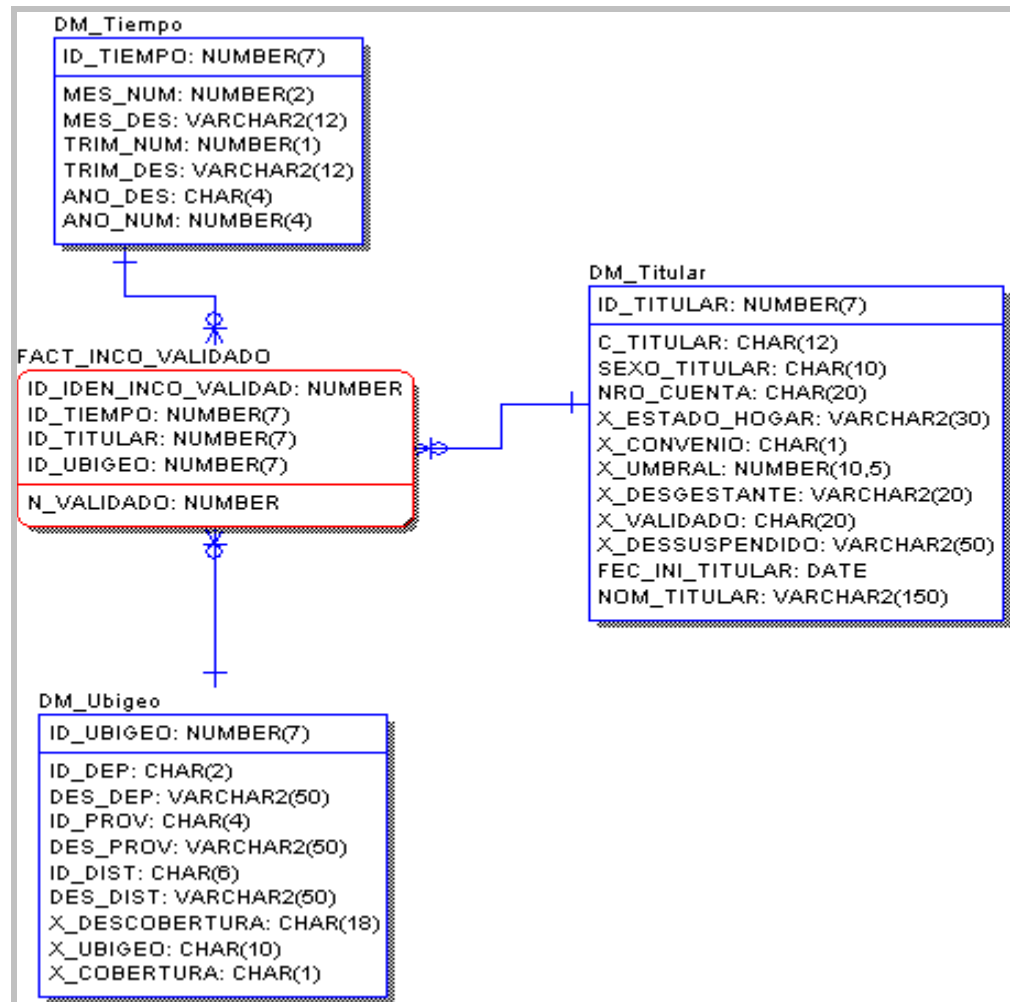


Figura 46: Modelo Estrella – Identificación de Hogares validados

3.3.12. Modelo Estrella: Identificación de Hogares Evaluados

Este modelo estrella describe las principales características de los titulares evaluados para la verificación del cumplimiento de compromisos (el hecho), analizada en base al ubigeo, etapa, sector (salud, educación) y el tiempo. La dimensión titular es incluida para obtener el detalle de los titulares evaluados. Con este modelo se pueden atender los siguientes requerimientos: REQCGF02, REQCGF05, REQCGF11, REQCGF12, REQCGF13, REQCGF14, REQCGF19, REQCGF01, REQCGF07, REQCGF20.

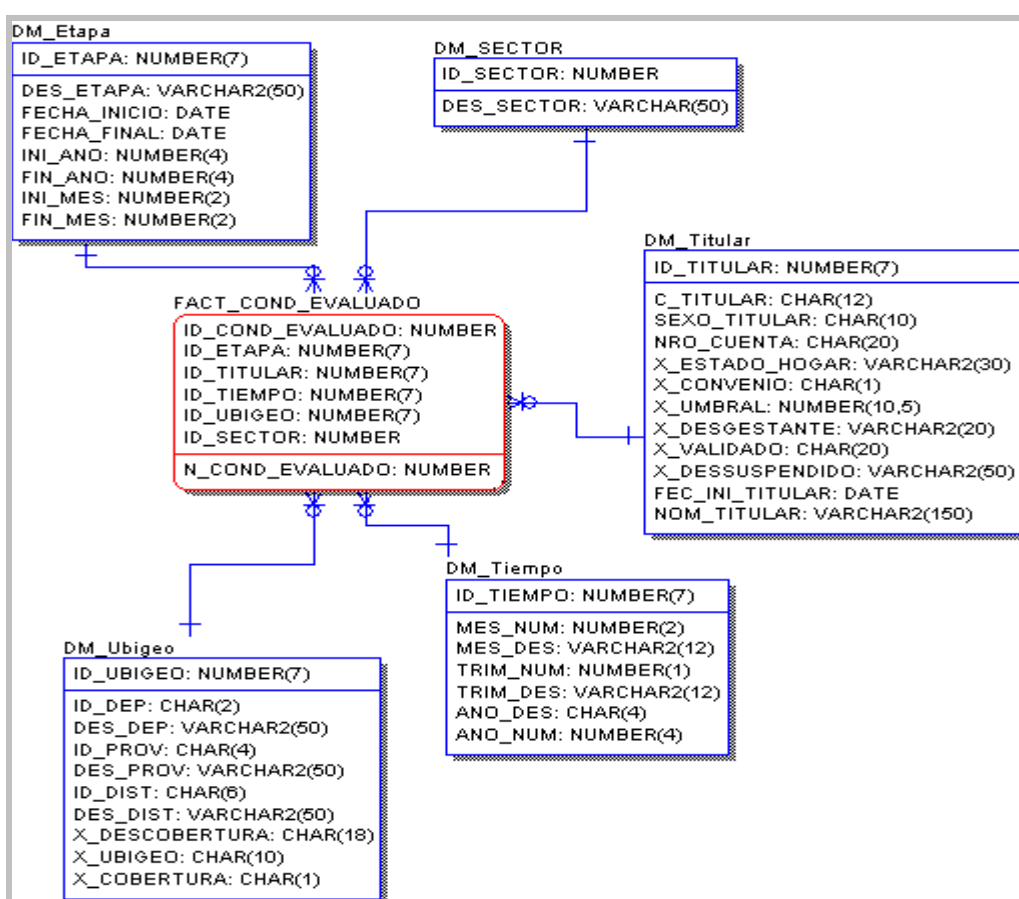


Figura 47: Modelo Estrella – Identificación de Hogares evaluados

3.3.13. Modelo Estrella: Identificación de hogares que incumplieron

Este modelo estrella describe las principales características de los titulares evaluados que incumplieron al menos uno de los compromisos (el hecho), analizada en base al ubigeo, etapa y el tiempo. La dimensión titular es incluida para obtener el detalle de los titulares evaluados. Con este modelo se pueden atender los siguientes requerimientos: REQCIG01, REQCIG02, REQCIG03, REQCIG04, REQCIG05, REQCIG06.

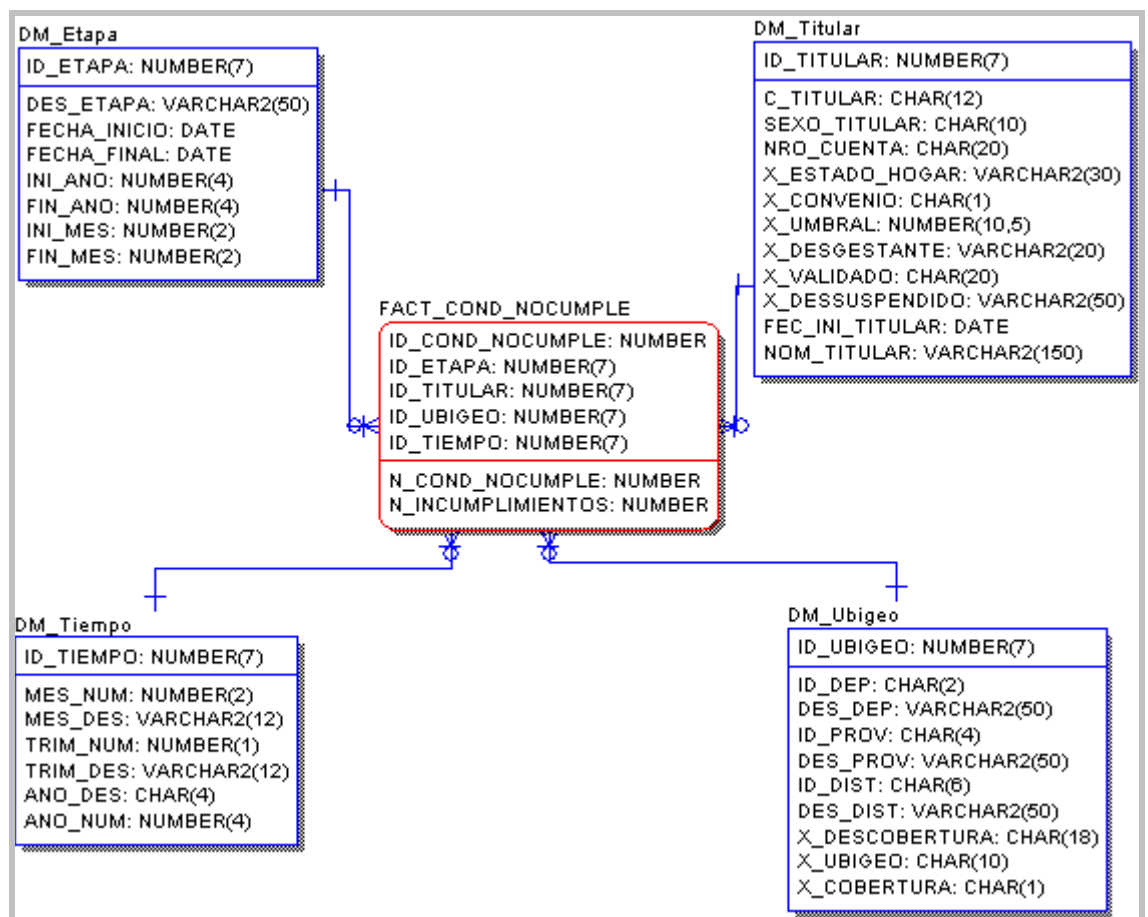


Figura 48: Modelo Estrella – Identificación de Hogares que incumplieron

3.3.14. Modelo Estrella: Identificación de niños que incumplieron en Educación

Este modelo estrella describe las principales características de los niños evaluados que incumplieron en el sector educación, en la verificación del cumplimiento de compromisos (el hecho), analizada en base al ubigeo, etapa, centro educativo y el tiempo. La dimensión beneficiario es incluida para obtener el detalle de los beneficiarios que incumplieron. Con este modelo se pueden atender los siguientes requerimientos: REQCEC04, REQCEC05, REQCEC06, REQCEC10, REQCEC11, REQCEC12.

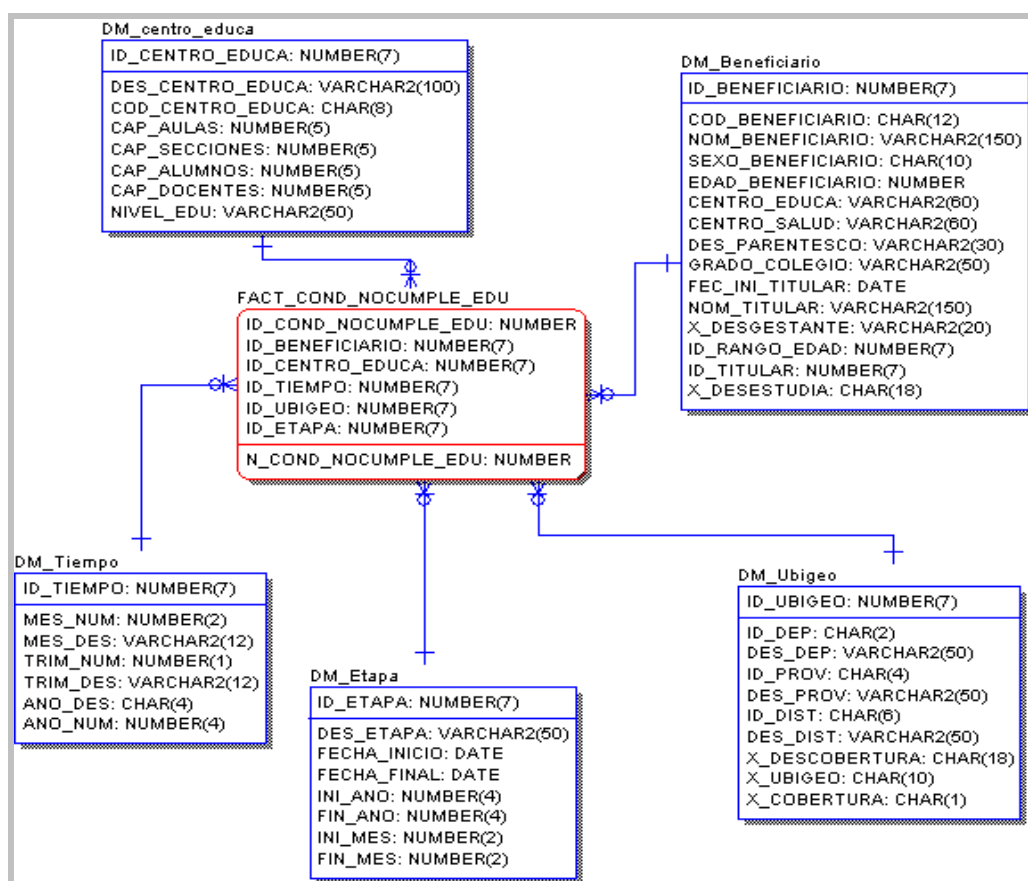


Figura 49: Modelo Estrella – Identificación de miembros que incumplieron en Educación

3.3.15. Modelo Estrella: Identificación de hogares que incumplieron en Salud

Este modelo estrella describe las principales características de los titulares evaluados que incumplieron en el sector salud, en la verificación del cumplimiento de compromisos (el hecho), analizada en base al ubigeo, etapa, centro de salud y el tiempo. La dimensión titular es incluida para obtener el detalle de los titulares evaluados. Con este modelo se pueden atender los siguientes requerimientos: REQCEC01, REQCEC02, REQCEC03.

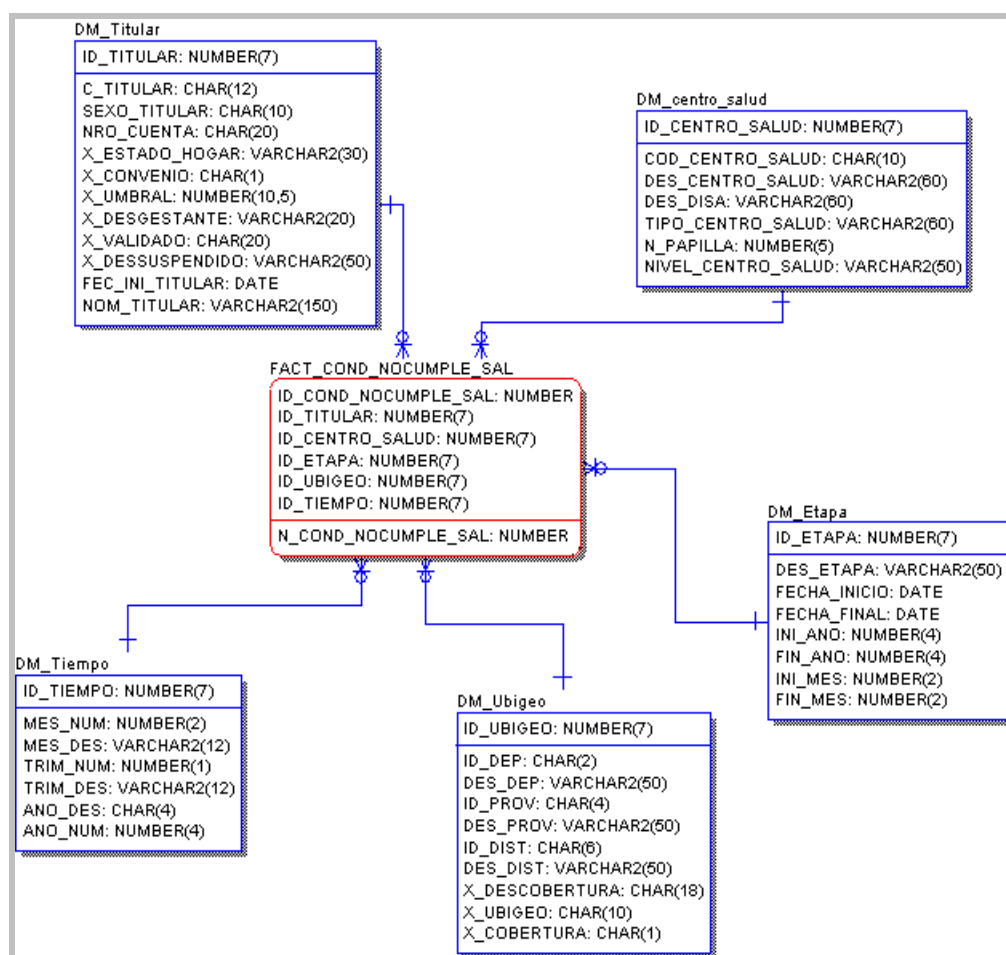


Figura 50: Modelo Estrella – Identificación de hogares que incumplieron en Salud

3.3.16. Modelo Estrella: Identificación de Titulares abonados

Este modelo estrella describe las principales características de los titulares abonados, es decir los que han recibido el incentivo económico (el hecho) analizada en base al ubigeo, la etapa, la modalidad de pago y el tiempo. La dimensión titular se incluye para obtener el detalle de los titulares abonados. . Con este modelo se pueden atender los siguientes requerimientos: REQPPP01, REQPPP02, REQPPP03, REQPPP04, REQPPP05, REQPPP06.

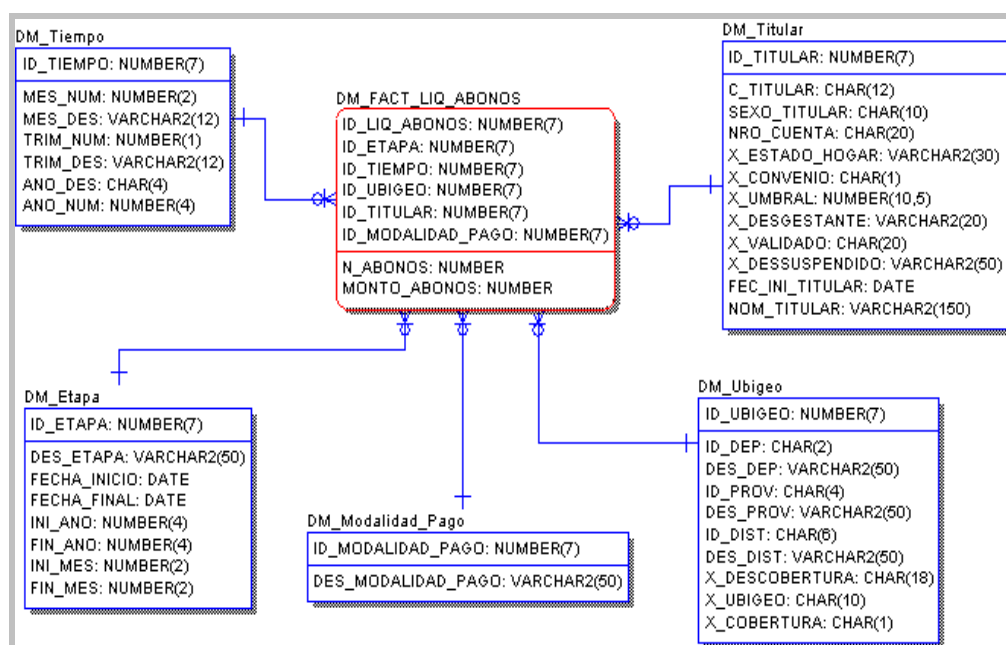


Figura 51: Modelo Estrella – Identificación de titulares abonados

3.3.17. Modelo Estrella: Identificación de Cuentas bancarias aperturadas

Este modelo estrella describe las principales características de las cuentas bancarias aperturadas para el PNADP (el hecho) analizada en base al ubigeo, la etapa y el tiempo. La dimensión cuenta bancaria se incluye para obtener el detalle de los cuentas bancarias aperturadas. Con este modelo se pueden atender los siguientes requerimientos: REQ PAC02, REQ PAC03, REQ PAC04, REQ PAC05.

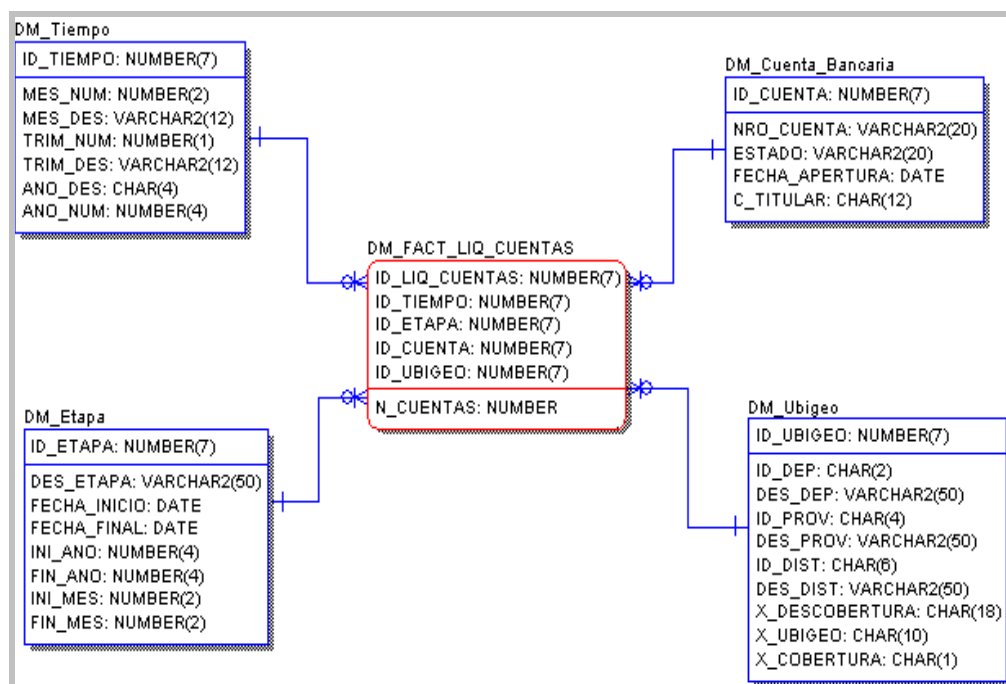


Figura 52: Modelo Estrella – Identificación de cuentas bancarias aperturadas

3.3.18. Modelo Estrella: Identificación de titulares del Padrón de Abonos

Este modelo estrella describe las principales características del padrón de pago (el hecho) analizada en base al ubigeo, la etapa, la modalidad de pago y el tiempo. La dimensión padrón se incluye para obtener el detalle del padrón de pago. Con este modelo se pueden atender los siguientes requerimientos: REQPPA01, REQPPA03, REQPPA07, REQPPA09, REQPPA11, REQACT13.

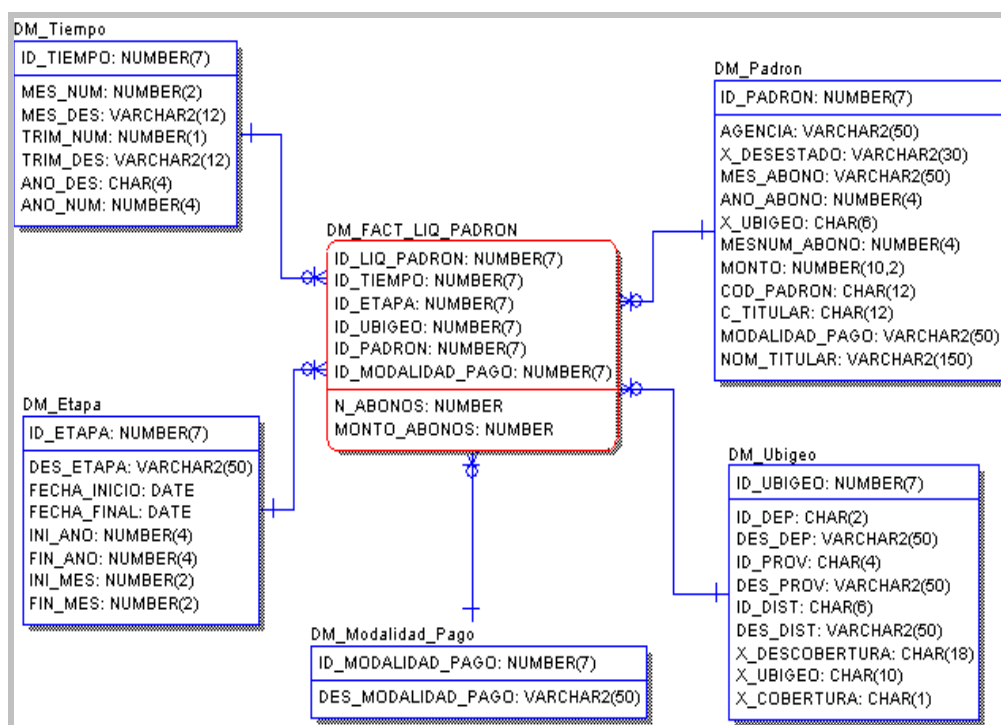


Figura 53: Modelo Estrella – Identificación de titulares del padrón de abonos

3.3.19. Modelo Estrella: Identificación de Titulares suspendidos

Este modelo estrella describe las principales características de los titulares suspendidos temporal o definitivamente (el hecho) analizada en base al ubigeo, la etapa, la modalidad de pago, el tipo de suspensión y el tiempo. La dimensión suspendidos se incluye para obtener el detalle de los titulares suspendidos. Con este modelo se pueden atender los siguientes requerimientos: REQAST01, REQAST02, REQAEX01, REQAEX04, REQACT01, REQACT02.

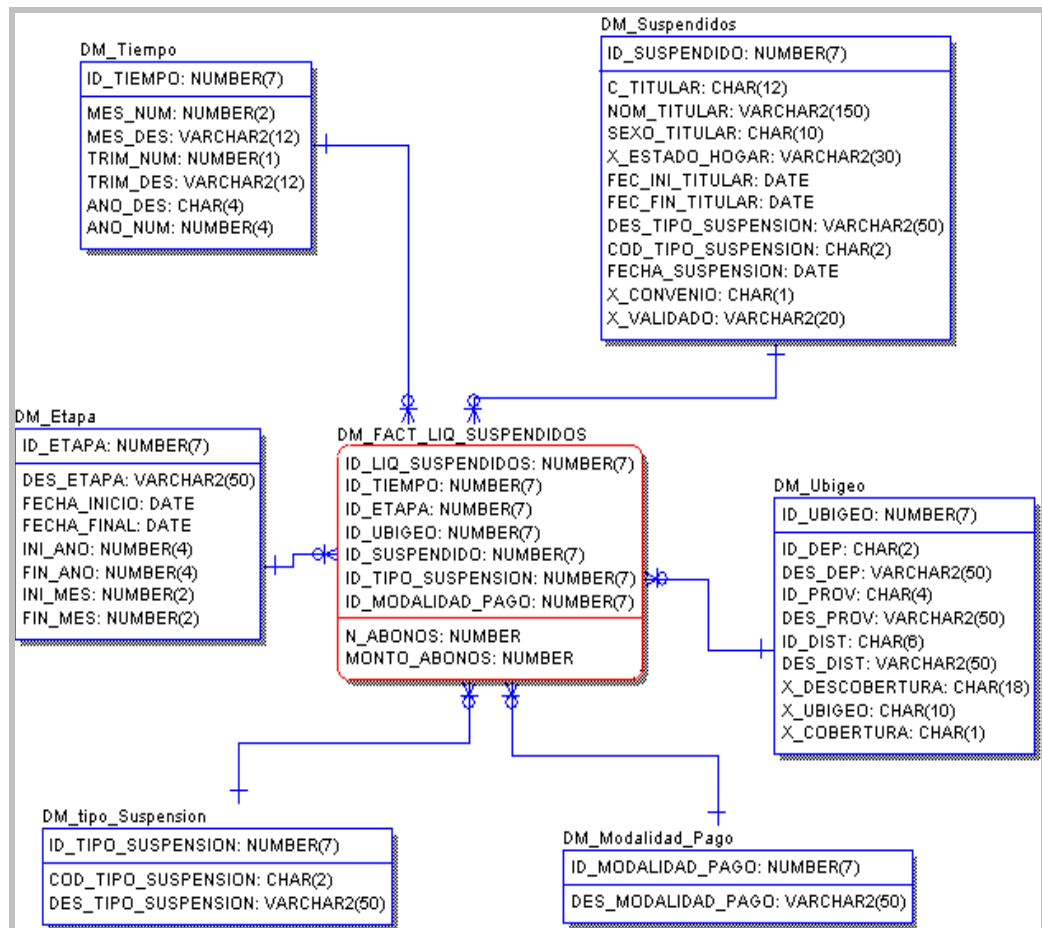


Figura 54: Modelo Estrella – Identificación de titulares suspendidos

3.3.20. Modelo Estrella: Identificación de Errores

Esta estrella es utilizada para el tratamiento de los errores y excepciones que se presenten durante los procesos de carga.

En este punto, se consideran ventanas que identificarán el tipo de error, la gravedad del mismo y la acción a tomar.

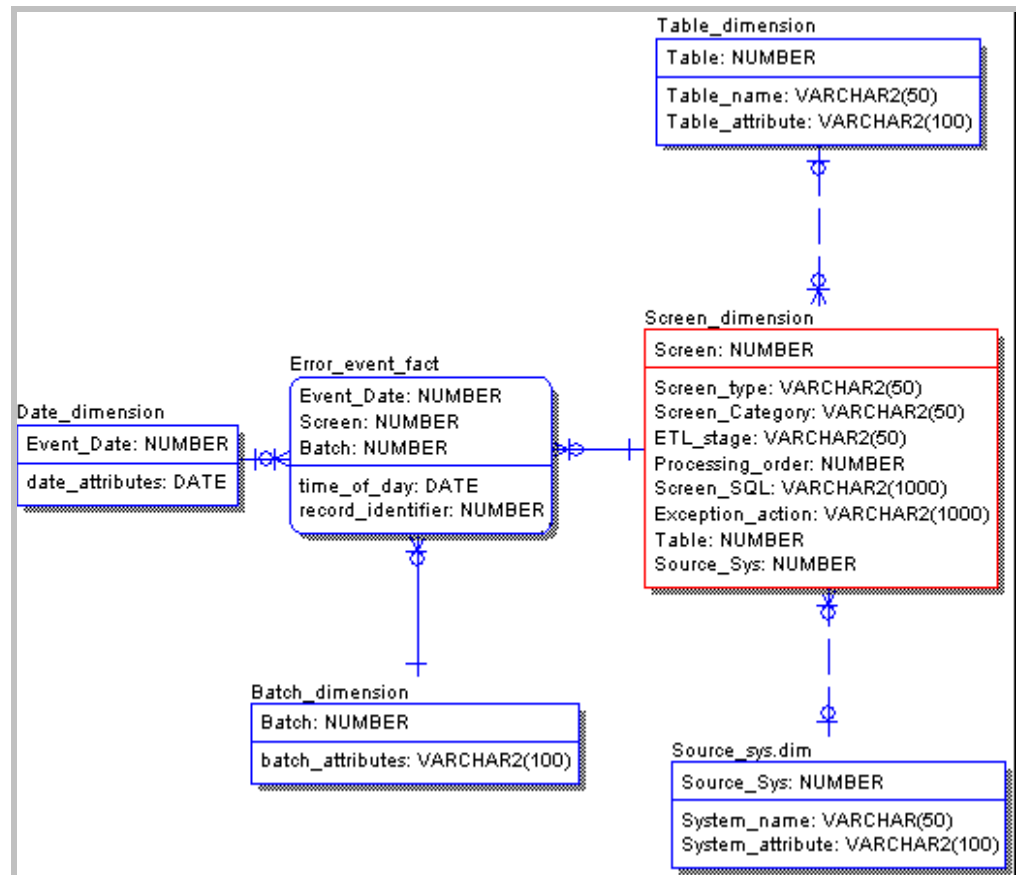


Figura 55: Modelo Estrella – Identificación de Errores

3.4 Arquitectura del Datawarehouse

La decisión de usar un *stage area* varia dependiendo del entorno y requerimientos de negocios. La capacidad de desarrollar un eficiente proceso ETL es parcialmente dependiente de la capacidad de determinar el balance adecuado entre lectura / escritura (I/O) física y procesamiento en memoria.

El reto de lograr este delicado balance entre escribir data en las tablas del *staging area* o mantenerla en memoria durante los procesos ETL, es una tarea que debe ser reconocida para poder crear procesos óptimos. El hecho que determina el usar o no un *stage area*, viene dado por dos objetivos contrapuestos:

- Llevar la data desde el origen hasta el Datawarehouse lo más rápido posible.
- Tener la capacidad de recuperarse desde cualquier error, sin necesidad de tener que reanudar todo el proceso.

Con esto en consideración, la arquitectura propuesta no considera el uso de un stage area, dado lo siguiente:

- Se dispone de un servidor dedicado para el Datawarehouse, con una capacidad de procesamiento y memoria adecuado.
(Ver sección 3.1.5.3)

- El tiempo estimado de la carga inicial y de las cargas incrementales, es lo suficiente corto para poder reanudar los procesos desde un inicio.
- Solo se hará uso de un origen de datos para el Datawarehouse, lo cual simplifica las labores de extracción, transformación y carga.
- Se requiere un diseño simple, ya que los tiempos de implementación y los esfuerzos para su posterior mantenimiento deben ser minimizados.

La figura 56 muestra el esquema simplificado de ETL, en este proceso se aprecia que no existe stage area.

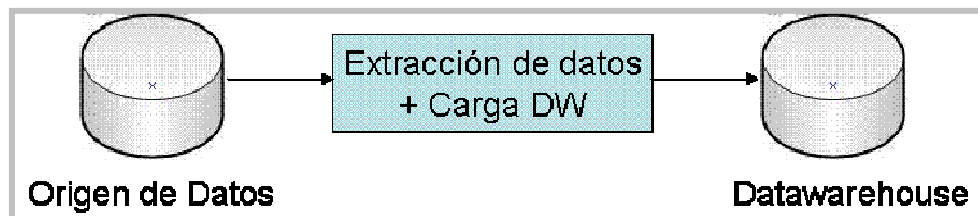


Figura 56: Proceso ETL

Con esto en consideración, la arquitectura del Datawarehouse es la que se muestra en la figura 57, en donde se aprecian las siguientes capas:

- La capa de datos fuente, con un único origen de datos.
- La capa de acceso a los datos, contempla los procesos ETL.

- La capa de Base de datos del Datawarehouse, la cual es administrada directamente por el sistema de administración de base de datos (RDBMS).
- La capa de acceso a la información, que puede ser accedida directamente o por medio de una herramienta de BI (BI Tool).
- La capa de aplicaciones, corresponde a las aplicaciones de escritorio, tableros de control, dashboards, etc.
- La capa de infraestructura de la metadata.

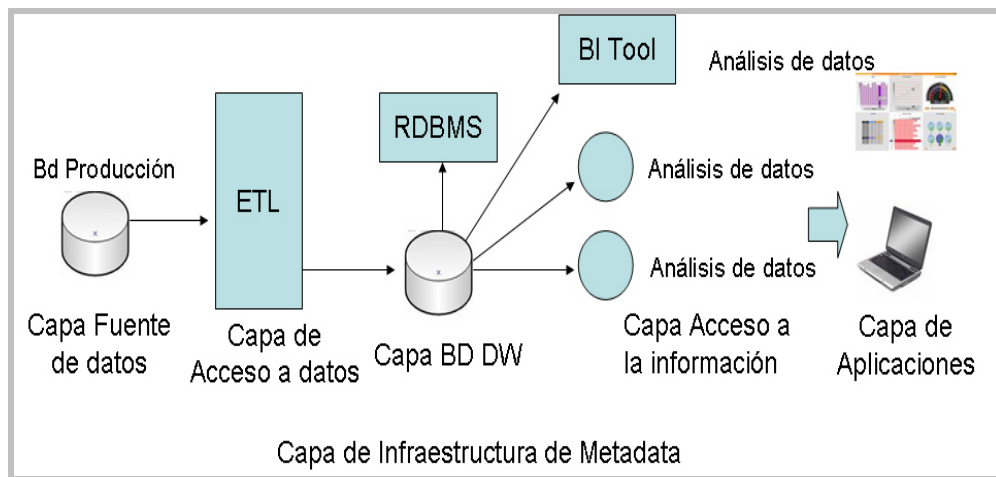


Figura 57: Arquitectura del Datawarehouse

3.5 Mapeo de Datos

El mapa lógico de datos será presentado como una tabla, incluyendo los siguientes campos:

- Nombre de la tabla origen, el nombre físico de la tabla que aparece en el Datawarehouse.
- Nombre de la columna origen, el nombre de la columna en la tabla del Datawarehouse.

- Tipo de tabla, indica si la tabla es una Fact table, dimensión o subdimensión.
- Base de datos origen, el nombre de la instancia de la base de datos donde reside la data origen.
- Tabla origen, el nombre de la tabla donde la data origen se encuentra, hay casos en los que varias tablas son requeridas.
- Columna origen, la columna o columnas necesarias para poblar el origen.
- Transformación, la manipulación requerida al origen de forma que corresponda al formato esperado en el destino.

Un ejemplo de la tabla de mapeo de datos se encuentra en el anexo 2.

3.6 Extracción de Datos y Transformación de datos

Como se indicó en la sección de Arquitectura de Datawarehouse, el diseño de la capa ETL no contempla un stage area, en vez de esto se hará uso de un modelo estrella para los aspectos de limpieza de datos. Se considerará el esquema Event error, propuesto en [8], y cuyo diseño se muestra en la figura 52.

A continuación se detallan los subprocesos ETL considerados en la solución.

3.6.1. Proceso ETL: Carga de información de censos

La carga de información de censos, es el proceso inicial donde se registra de forma masiva a las viviendas censadas y población censada total, pobre y no pobre. El proceso se inicia con la extracción de los datos de las tablas tpoblacion, tetapa y tvivienda. Las transformaciones se aplican a estas tres estructuras y finalmente el proceso de carga a las DIM y Fact tables, donde los registros rechazados se almacenan en el esquema Event Error.

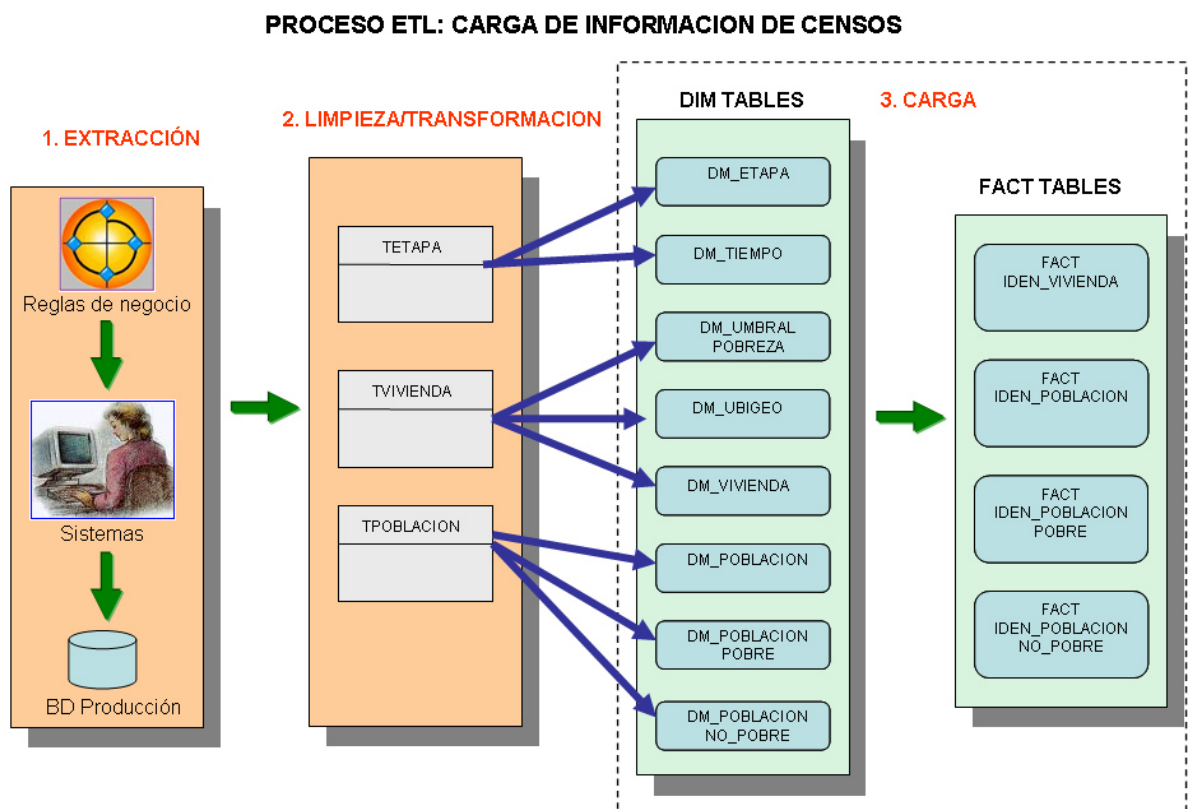


Figura 58: Proceso ETL - Carga de información de censos

3.6.2. Proceso ETL: Carga de información de condicionalidades

La carga de información de evaluación y cumplimiento de compromisos, es el proceso donde se certifica la asistencia de los beneficiarios a los sectores de salud y educación. El proceso se inicia con la extracción de los datos de los beneficiarios y de los sectores. Las transformaciones se aplican a estas estructuras y finalmente el proceso de carga, donde los registros rechazados se almacenan en el esquema Event Error.

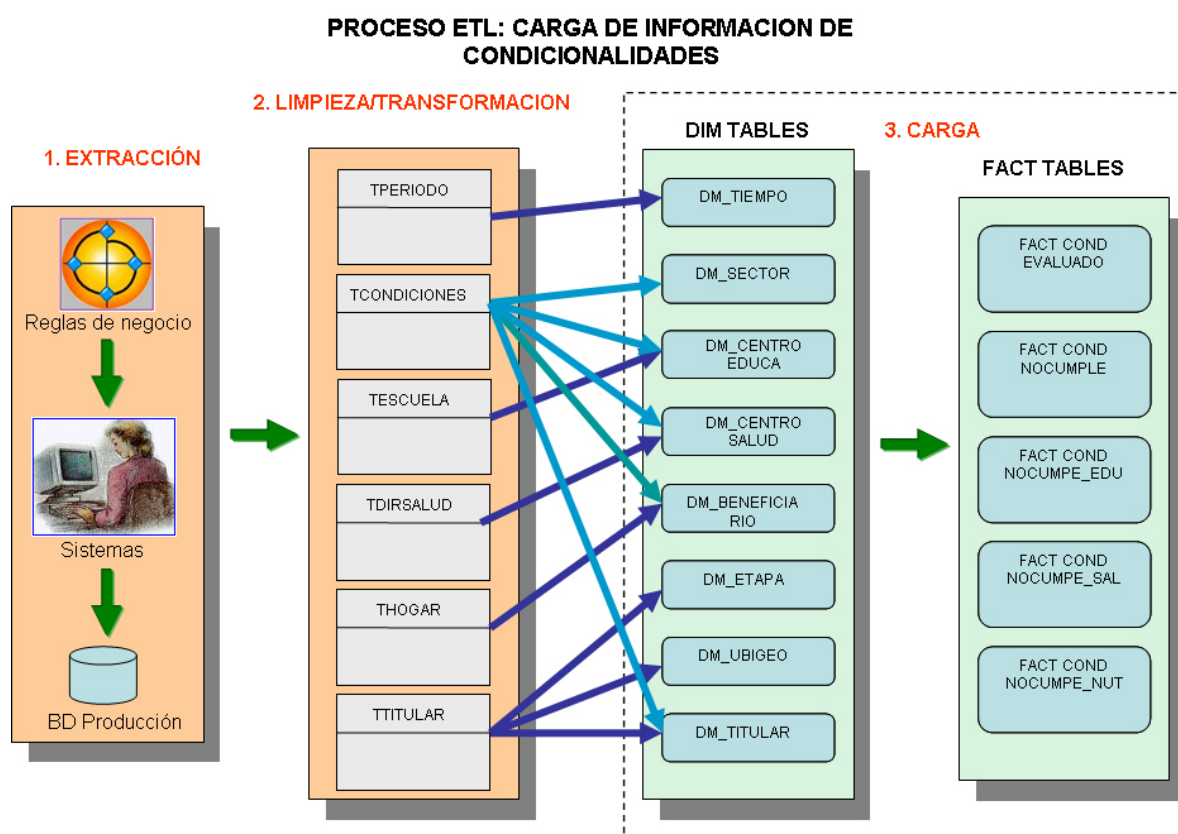


Figura 59: Proceso ETL - Carga de información de condicionalidades

3.6.3. Proceso ETL: Carga de información del padrón de abonos

La carga de información del padrón de abonos, es el proceso donde se determinará los hogares beneficiados que recibirán el incentivo económico por mes. El subproceso se inicia con la extracción de los datos de los beneficiarios, puntos de pago y modalidad de pago. Las transformaciones se aplican a estas y finalmente el proceso de carga, donde los registros rechazados se almacenan en el esquema Event Error.

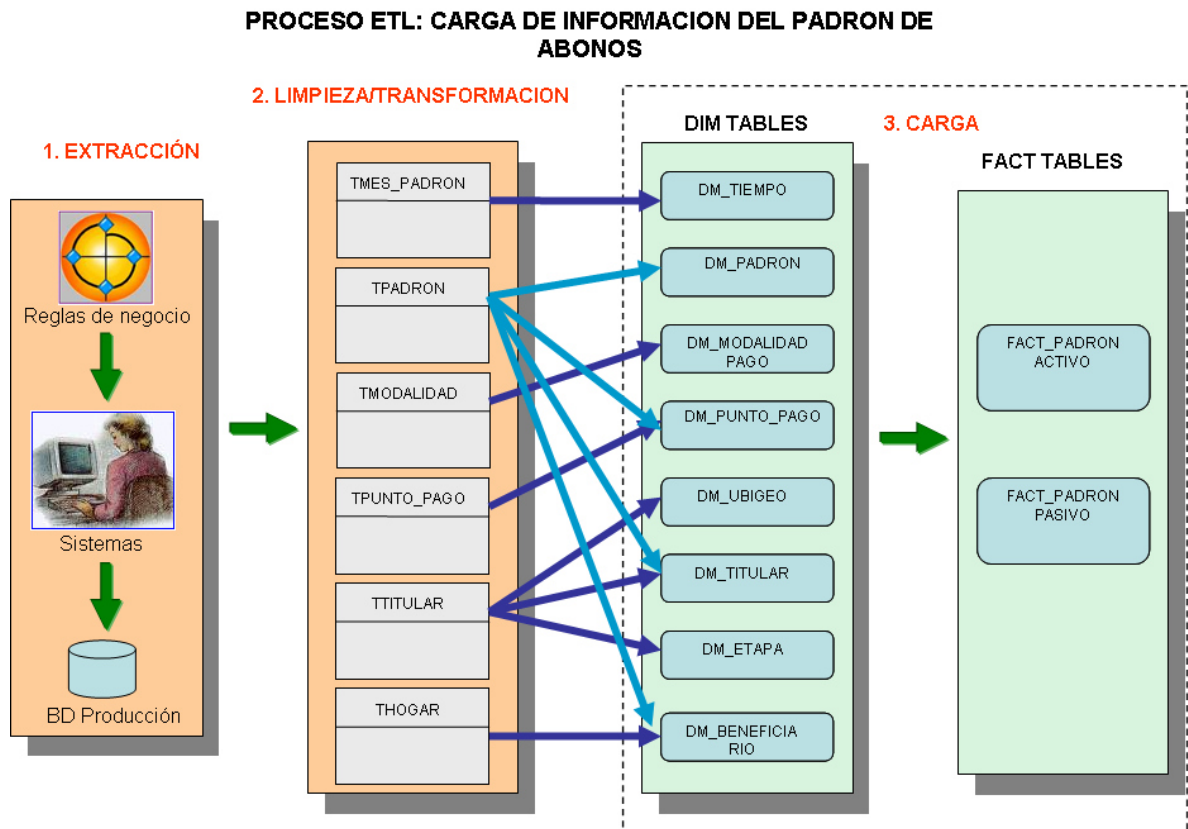


Figura 60: Proceso ETL - Carga de información del padrón de abonos

3.6.4. Proceso ETL: Carga de información histórica mensual

La carga de información histórica mensual, es el proceso por el cual se almacenará los históricos de las principales tablas por mes. El proceso se inicia con la extracción de los datos de las tablas históricas de titular, beneficiarios y suspendidos. Las transformaciones se aplican a estas estructuras, básicamente agrupaciones y sumalizaciones. Finalmente el proceso de carga, donde los registros rechazados se almacenan en el esquema Event Error.

PROCESO ETL: CARGA DE INFORMACION HISTORICA MENSUAL

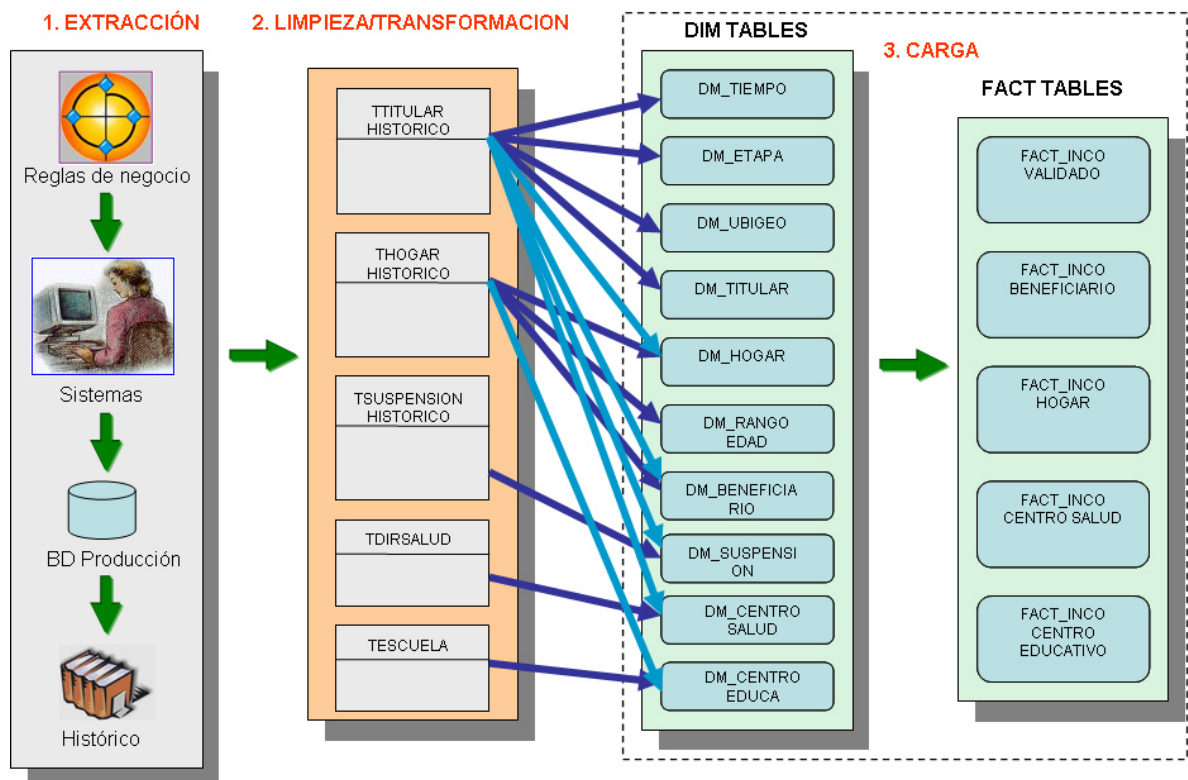


Figura 61: Proceso ETL - Carga de información histórica mensual

Para la extracción de los datos se han realizado procedimientos almacenados, los cuales leen los datos del esquema de la base de datos Stand By. La lectura se hace sobre la base de datos stand by teniendo en consideración, que esta se refresca diariamente. La transformación se realiza en el mismo entorno de extracción. Los procedimientos se han agrupado en paquetes, bajo el siguiente esquema:

- PACK_CARGA_INICIAL, para los procedimientos de carga inicial. Existe un procedimiento para cada dimensión y cada Fact. table.
- PACK_CARGA_INCREMENTAL, para los procedimientos de carga incremental. Existe un procedimiento para cada dimensión y cada Fact table.

3.7 Carga del Datawarehouse

Se han definido dos mecanismos de carga: carga inicial y carga incremental. La carga Inicial realiza la carga completa del Datawarehouse. Se ejecuta una sola vez. La carga Incremental se basa en un mecanismo de triggers, los cuales marcan los registros que han sufrido alguna clase de actualización, para posteriormente ser incluidos en los procedimientos de actualización del Datawarehouse. Se ejecuta diariamente.

Tiempos estimados de carga:

- Carga Inicial: aproximadamente 2 horas y 10 min. a 3 horas.
- Carga Incremental: aproximadamente 5 min. a 30 min.

Previo a la carga inicial, se iniciará el testeo formal del mismo, el cual considerará los siguientes parámetros y métricas de prueba:

Métrica	Valor deseable	Umbral
Tiempos de carga inicial.	2 horas	No mayor a 3 horas
Tiempos de carga incrementales.	10 minutos	No mayor a 30 minutos.
Registros rechazados en el proceso de transformación.	0%	No mayor al 0.5 %
Recursos consumidos	50% de la capacidad de procesamiento del servidor.	No mayor al 80% en las transacciones mas “pesadas”
Número de reinicios del proceso de carga.	0	0

Figura 62: Parámetros y métricas de prueba

Las unidades de testeo serán cada tabla del modelo dimensional (“Fact table” y Dimensiones). No se testearán los registros que no han pasado el proceso de transformación. Los testeos se realizarán sobre el entorno stand by.

La integración de los testeos contemplará lo siguiente:

- La secuencia de los procedimientos ETL.
- La carga inicial de los registros del Datawarehouse.
- La carga incremental, de los registros que han sufrido actualizados en la base de datos operacional.
- Los logs de generación de errores.

Escenarios

Los escenarios de validación serán:

- Validación de la cantidad de registros cargados, para esto se hará la comparación en base a consultas obtenidas de la base de datos operacional (procesos de aislamiento).
- Validación dimensional, se validará la integridad entre las tablas y sus relaciones.
- Análisis estadístico, acá se validan diversos cálculos y resúmenes.
- Validación de la calidad de datos, se revisará los datos faltantes o perdidos, y los datos inconsistentes.
- Granularidad, se validará hasta el máximo nivel de granularidad posible, en nuestro caso, hasta el nivel de distrito.

3.8 Liberación de la metadata

Se consideran los siguientes tipos de Metada:

- Metadata del negocio (Ver sección 2.3)
- Metadata técnica (Ver Anexo 1)

3.9 Plan de Prueba y validación del Datawarehouse

En la figura 63 se muestra las métricas que se han definido para la validación del Datawarehouse.

Métrica	Valor deseable	Umbral
Tiempo de respuesta	10 s.	No mayor a 30 s.
Percepción de facilidad de uso. Tiempo en entender la herramienta.	30 min.	No mayor a dos horas.
Percepción de usabilidad.	100%	Al menos ante el 80% de los requerimientos de análisis de información.
Número de pasos para obtener la información deseada.	3	No mayor a 5 pasos.
Nivel de disponibilidad	24 x 7	De lunes a domingo, salvos horas de mantenimiento programados.
Nivel de confiabilidad	100%	100%

Figura 63: Parámetros y métricas de validación del Datawarehouse

Unidades de testeo

Las unidades de testeo corresponderán a cada modelo dimensional definidos. Todas las pruebas serán pruebas funcionales con los usuarios, no se realizarán testeos automatizados. Los testeos se realizarán sobre el entorno stand by. Para la validación funcional se hará uso del ORACLE discoverer.

Aceptación de usuarios.

Tanto la formulación de requerimientos como la aceptación de los mismos se harán en concordancia al “Procedimiento de control de cambios a los sistemas de información”, contenido en el plan de seguridad del PNADP. Ver Anexo 3.

Finalmente, concluidas las pruebas técnicas y funcionales se realizará el pase al entorno de producción del Datawarehouse. Este procedimiento se hará en concordancia al “Procedimiento control de programas antes de su pase a producción”, contenido en el plan de seguridad del PNADP. Ver Anexo 4.

CAPITULO IV

4. CONCLUSIONES Y TRABAJOS FUTUROS

4.1 Conclusiones

- La administración de un proyecto de Datawarehouse, involucra diferentes aspectos, adicionales al diseño de la solución. Estos aspectos, entre otros, son: La estrategia del Datawarehouse, los procesos administrativos de soporte, etc.
- Los usuarios deben estar profundamente comprometidos en los esfuerzos de implementación de Datawarehouse. En este punto los beneficios de la inteligencia de negocios deben ser comprendidos por los usuarios clave antes del inicio del proyecto.
- El aspecto de calidad de datos es trascendental en todo esfuerzo de implementación de un Datawarehouse. Este punto no debe ser trivializado, y de ser el caso, debe ser considerado como un proyecto previo.
- La oferta de soluciones en el campo de inteligencia de negocios y de Datawarehouse en particular, es bastante amplia y en constante cambio. Por lo tanto, la orientación de la implementación debe hacerse más hacia la metodología que hacia al aspecto tecnológico.
- Las soluciones de inteligencia de negocios son también aplicables y generadoras de valor para entidades gubernamentales - sin fines de lucro.

- Si bien existen muchas metodologías para la implementación de un Datawarehouse, éstas comparten aspectos en común: análisis de requerimientos del negocio, diseño de datos, diseño de arquitectura, desarrollo en implementación de la solución.

4.2 Recomendaciones

- Dada la trascendencia y particularidad de la solución planteada, es necesaria la implementación de una política de Administración del Datawarehouse, la cual debe ser incluida en el Plan de seguridad de información del PNADP.
- El modelo planteado, esta fuertemente orientado a los procesos actuales del PNADP, por lo tanto su revisión constante debe ser incluida como parte de la política propuesta en el punto anterior.
- Instruir al staff de la Gerencia de sistemas en el plan de Datawarehouse y en la política planteada para la administración del mismo.
- Se deben realizar talleres de capacitación para los usuarios finales, en dos niveles, dirigidos a la alta gerencia (incluirán temas como tableros de mando, BSC), y al nivel táctico operacional (incluirá temas como utilización de herramientas de BI).
- Los procedimientos de respaldo y backup del Datawarehouse deben realizarse en concordancia al procedimiento “Respaldos de información y backups”, incluido en el plan de seguridad del PNADP.

- Se debe considerar una estrategia diferenciada de implementación del Datawarehouse, una para el nivel gerencial, y otra para el nivel táctico-operacional.
- La estrategia de implementación del Datawarehouse en el PNADP debe ser down-top (abajo hacia arriba).

4.3 Trabajos Futuros

Este trabajo servirá como base para los sucesivos proyectos de desarrollo de aplicaciones:

- Proyecto tablero de mando para el PNADP. El análisis y diseño de un balanced score card, haciendo uso de la infraestructura existente.
- Diseño de reportes gerenciales. Este proyecto involucra determinar cuales son los principales cuadros resúmenes requeridos por la alta dirección, en base a los informes de avance de la gestión requeridos por la Presidencia del Consejo de Ministros (PCM).
- Diseño de un datamining² para el PNADP. Haciendo uso de algoritmos avanzados de redes neuronales, para determinar los patrones que describen los principales aspectos de las incorporaciones, evaluación de condicionalidades y pagos.

² Se denomina data mining al conjunto de técnicas encaminadas a la extracción de "conocimiento" procesable implícito en las bases de datos de las empresas. Las bases de la minería de datos se encuentran en la inteligencia artificial y en el análisis estadístico.

BIBLIOGRAFIA

- [1] Beyond Datawarehousing: What's Next in Business Intelligence?.
Matteo Golfarelli
DEIS - University of Bologna, Stefano Rizzi DEIS - University of Bologna,
Iuris Cella Gruppo Formula S.p.A.
- [2] Modeling, Querying and Reasoning about OLAP Databases: A
Functional Approach. Ken Q. Pu University of Toronto Toronto, ON, Canada.
- [3] Towards OLAP Security Design – Survey and Research Issues Torsten
Priebe, Günther Pernul Department of Information Systems University of Essen
Universitätsstr. 9, 45141 Essen Germany.
- [4] MODELING STRATEGIES AND ALTERNATIVES FOR DATA
WAREHOUSING PROJECTS. Nenad Jukic. Communications of the ACM. April
2006.
- [5] Tendencias en Análisis de Alto rendimiento Yossi Matias. HyperRoll
Inc. & Stanford University.
- [6] Moss, L. And Atre, S. Business Intelligence Roadmap. Addison-
Wesley. 2003.
- [7] Imhoff, C., Gallemmo, N., and Geiger, J. Mastering Data Warehouse
Design: Relational and Dimensional Techniques. Wiley, 2003.
- [8] The Data Warehouse ETL Toolkit. Ralph Kimball. Wiley Publishing Inc.
2004.
- [9] Enhancing data quality in Datawarehouse environments.
Communications of the ACM. January 1999.
- [10] A comparison of Data Warehouse Methodologies. Arun Sen y Atish
P. Sinha. Communications of the ACM. March 2005.

[11] A typology of data warehouse quality. Cindy Le Rouge, University of South of Florida. Creggan Gjestland, University of Alabama. 2002 – Eight Americas Conference on Information Systems.

[12] Datawarehousing. La integración de información para la mejor toma de decisiones. Harjinder S. Gill, Prakash C. Rao. Prentice Hall 1996.

[13] An Exploratory Investigation of System Success Factors in Datawarehousing. Bongsik Shin. Department of Information and Decision Systems. San Diego State University. Journal of the Association for Information Systems 2003.

GLOSARIO

Datawarehouse. Es una colección de datos orientadas a un dominio, integrado, no volátil y variable en el tiempo que ayuda a la toma de decisiones de la empresa u organización. Se trata, sobre todo, de un expediente de una empresa más allá de la información transaccional y operacional, almacenada en una base de datos diseñada para favorecer análisis y la divulgación eficientes de datos.

Datamart. Un datamart es una versión especializada de un datawarehouse. Al igual que un datawarehouse, un datamart contiene una “fotografía” de la data operacional que ayuda a la empresa a la estrategia, en ase al análisis de tendencias y experiencias. La principal diferencia es que la creación de un datamart esta orientada a una especifica, predefinida necesidad de cierto predefinido y agrupado conjunto de datos

Inteligencia de Negocios. El conjunto de estrategias y herramientas enfocadas a la administración y creación de conocimiento mediante el análisis de datos existentes en una organización o empresa.

OLAP. Es el acrónimo en inglés de procesamiento analítico en línea (On-Line Analytical Processing). OLAP es un tipo de tecnología que permite a los usuarios mejorar la visión que tienen de sus datos de manera rápida, interactiva y fácil de usar.

MOLAP. (multidimensional online analytical processing – procesamiento analítico multidimensional en línea), la premisa es el procesamiento

transaccional interactivo (OLTP) está mejor implementado almacenando los datos de forma multidimensional.

ROLAP. Relational Online Analytical Processing, o proceso analítico relacional en línea. Implica guardar los datos en un modelo relacional; en este caso se proporcionan mejor las características OLAP usando una base de datos relacional.

Modelo Entidad Relacion – MER. Los diagramas o modelos entidad-relación (a veces denominado por su siglas, E-R "Entity relationship") son una herramienta para el modelado de datos de un sistema de información. Estos modelos expresan entidades relevantes para un sistema de información, sus inter-relaciones y propiedades.

Dashboard. Un conjunto pequeño y definido de métricas claves usadas para proveer una rápida evaluación de un proyecto o del estado de un proceso.

Diccionario de datos

Tablas

Nombre
DM_BENEFICIARIO
DM_CENTRO_EDUCA
DM_CENTRO_SALUD
DM_CUENTA_BANCARIA
DM_ETAPA
DM_FACT_IDEN_HOGAR_ELEGIBLE
DM_FACT_IDEN_HOGAR_OBJETIVO
DM_FACT_IDEN_HOGAR_TOTAL
DM_FACT_IDEN_HOGAR_VALIDADO
DM_FACT_IDEN_POBLACION
DM_FACT_IDEN_POBLACION_POBRE
DM_FACT_IDEN_VIVIENDA
DM_FACT_INCO_BENEFICIARIO
DM_FACT_INCO_CENTRO_EDUCA
DM_FACT_INCO_CENTRO_SALUD
DM_FACT_INCO_HOGAR
DM_FACT_INCO_PROMOTOR
DM_FACT_INCO_PROMOTOR_DEP
DM_FACT_LIQ_ABONOS
DM_FACT_LIQ_CUENTAS
DM_FACT_LIQ_PADRON
DM_FACT_LIQ_SUSPENDIDOS
DM_HOGAR
DM_HOGAR_ELEGIBLE
DM_HOGAR_OBJETIVO
DM_HOGAR_TOTAL
DM_HOGAR_VALIDADO
DM_MODALIDAD_PAGO
DM_PADRON
DM_POBLACION
DM_POBLACION_POBRE
DM_PROMOTOR
DM_RANGO_EDAD
DM_SUSPENDIDOS
DM_TIEMPO
DM_TIPO_PROMOTOR
DM_TIPO_SUSPENSION
DM_TITULAR
DM_UBIGEO
DM_UBIGEO_DEP
DM_UMBRAL_POBREZA
DM_VIVIENDA

Columna(s) de "DM_BENEFICIARIO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_BENEFICIARIO	NUMBER(7)	NOT NULL	Código de identificación único del beneficiario	Yes
COD_BENEFICIARIO	CHAR(12)	NULL	Almacena el código de población único del esquema transaccional del PNADP.	No
NOM_BENEFICIARIO	VARCHAR2(150)	NULL	Nombre de la persona	No
SEXO_BENEFICIARIO	CHAR(10)	NULL	Sexo de la persona	No
EDAD_BENEFICIARIO	NUMBER	NULL	Edad de la persona	No
CENTRO_EDUCA	VARCHAR2(100)	NULL	Centro de Educación donde se estudia el beneficiario	No
CENTRO_SALUD	VARCHAR2(60)	NULL	Centro de Salud donde se atiende el beneficiario	No
DES_PARENTESCO	VARCHAR2(30)	NULL	Parentesco del beneficiario con su	No

Columna(s) de "DM_BENEFICIARIO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			representante o titular.	
GRADO_COLEGIO	VARCHAR2(50)	NULL	Grado actual de colegio del beneficiario.	No
FEC_INI_TITULAR	DATE	NULL	Fecha de inicio del Titular, obtenido del esquema transaccional de la tabla titular del campo f_inicio	No
NOM_TITULAR	VARCHAR2(150)	NULL	Almacena el Nombre del titular del Beneficiario	No
X_DESGESTANTE	VARCHAR2(20)	NULL	Descripción del estado gestante	No
ID_RANGO_EDAD	NUMBER(7)	NULL	Código de identificación del rango de edad al cuál pertenece el beneficiario, se enlaza con la dimensión DM_RANGO_EDAD	No
ID_TITULAR	NUMBER(7)	NULL	Código de identificación del Titular del beneficiario, se enlaza con la dimensión DM_TITULAR como un copo de nieve.	No
X_DESESTUDIA	CHAR(18)	NULL	Descripción de estado estudiante	No

Columna(s) de "DM_CENTRO_EDUCA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
DES_CENTRO_EDUCA	VARCHAR2(100)	NULL	Descripción del Centro educativo	No
ID_CENTRO_EDUCA	NUMBER(7)	NOT NULL	Código de identificación único del centro educativo	Yes
COD_CENTRO_EDUCA	CHAR(8)	NULL	Código modular del Centro Educativo	No
CAP_AULAS	NUMBER(5)	NULL	Capacidad de aulas	No
CAP_SECCIONES	NUMBER(5)	NULL	Capacidad de secciones	No
CAP_ALUMNOS	NUMBER(5)	NULL	Capacidad de Alumnos	No
CAP_DOCENTES	NUMBER(5)	NULL	Capacidad de docentes	No
NIVEL_EDU	VARCHAR2(50)	NULL	Nivel educativo máximo otorgado	No

Columna(s) de "DM_CENTRO_SALUD" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_CENTRO_SALUD	NUMBER(7)	NOT NULL	Código de identificación único del centro de salud	Yes
COD_CENTRO_SALUD	CHAR(10)	NULL	Código del centro de salud	No
DES_CENTRO_SALUD	VARCHAR2(60)	NULL	Descripción del centro de salud	No
DES_DISA	VARCHAR2(60)	NULL	División de salud a la que pertenece el centro de salud	No
TIPO_CENTRO_SALUD	VARCHAR2(60)	NULL	Descripción : centro o puesto	No
N_PAPILLA	NUMBER(5)	NULL		No
NIVEL_CENTRO_SALUD	VARCHAR2(50)	NULL	Descripción : centros o puestos	No

Columna(s) de "DM_CUENTA_BANCARIA" Tabla				
Name	Datatype	Null Option	Comment	Is PK

Columna(s) de "DM_CUENTA_BANCARIA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_CUENTA	NUMBER(7)	NOT NULL	Código de identificación único de la Cuenta Bancaria.	Yes
NRO_CUENTA	VARCHAR2(20)	NULL	Número de Cuenta Bancaria, se obtiene del esquema transaccional del PNADP.	No
ESTADO	VARCHAR2(20)	NULL	Estado de la cuenta bancaria	No
FECHA_APERTURA	DATE	NULL	Fecha de apertura de la cuenta bancaria	No
C_TITULAR	CHAR(12)	NULL	Almacena el código de población único del esquema transaccional del PNADP.	No

Columna(s) de "DM_ETAPA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_ETAPA	NUMBER(7)	NOT NULL	Código identificador único de etapa.	Yes
DES_ETAPA	VARCHAR2(50)	NULL	Descripción de la etapa, ejemplo: Etapa1.	No
FECHA_INICIO	DATE	NULL	Fecha de inicio de la etapa.	No
FECHA_FINAL	DATE	NULL	Fecha final de la etapa.	No
INI_ANO	NUMBER(4)	NULL	Año de inicio.	No
FIN_ANO	NUMBER(4)	NULL	Año de fin.	No
INI_MES	NUMBER(2)	NULL	Mes de inicio.	No
FIN_MES	NUMBER(2)	NULL	Mes de fin.	No

Columna(s) de "DM_FACT_IDEN_HOGAR_ELEGIBLE" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_HOGAR_ELEGIBLE	NUMBER(7)	NOT NULL	Código identificador único de la FactTabla de DM_FACT_IDEN_HOGAR_ELEGIBLE	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código identificador del Rango de índice de pobreza se enlaza con la dimensión DM_UBIGEO.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código identificador de Tiempo se enlaza con la dimensión DM_TIEMPO.	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código identificador de Etapa se enlaza con la dimensión DM_ETAPA.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código identificador del Ubigeo se enlaza con la dimensión DM_UBIGEO.	Yes
N_HOGAR_ELEGIBLE	NUMBER	NULL	Cantidad de Hogares pobres elegibles	No
ID_HOGAR_ELEGIBLE	NUMBER(7)	NOT NULL	Código identificador del Hogar Elegible se enlaza con la dimensión DM_HOGAR_ELEGIBLE.	Yes

Columna(s) de "DM_FACT_IDEN_HOGAR_OBJETIVO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_HOGAR_OBJ	NUMBER(7)	NOT NULL	Código identificador	Yes

Columna(s) de "DM_FACT_IDEN_HOGAR_OBJETIVO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ETIVO			único de la FactTabla de DM_FACT_IDEN_HOGAR_OBJETIVO.	
ID_UMBRAL	NUMBER(7)	NOT NULL	Código identificador del Rango de índice de pobreza se enlaza con la dimensión DM_UBIGEO.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código identificador de Tiempo se enlaza con la dimensión DM_TIEMPO.	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código identificador de Etapa se enlaza con la dimensión DM_ETAPA.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código identificador del Ubigeo se enlaza con la dimensión DM_UBIGEO.	Yes
N_HOGAR_OBJETIVO	NUMBER	NULL	Cantidad de Hogares objetivo	No
ID_HOGAR_OBJETIVO	NUMBER(7)	NOT NULL	Código identificador del Hogar Objetivo se enlaza con la dimensión DM_HOGAR_OBJETIVO.	Yes

Columna(s) de "DM_FACT_IDEN_HOGAR_TOTAL" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_HOGAR_TOTAL	NUMBER(7)	NOT NULL	Código identificador único de la FactTabla de DM_FACT_IDEN_HOGAR_TOTAL.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código identificador de Tiempo se enlaza con la dimensión DM_TIEMPO.	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código identificador del Rango de índice de pobreza se enlaza con la dimensión DM_UBIGEO.	Yes
ID_HOGAR_TOTAL	NUMBER(7)	NOT NULL	Código identificador del Hogar censado se enlaza con la dimensión DM_HOGAR_TOTAL.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código identificador del Ubigeo se enlaza con la dimensión DM_UBIGEO.	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código identificador de Etapa se enlaza con la dimensión DM_ETAPA.	Yes
N_HOGAR_TOTAL	NUMBER	NULL	Cantidad de Hogares Censados	No

Columna(s) de "DM_FACT_IDEN_HOGAR_VALIDADO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_HOGAR_VALIDADO	NUMBER(7)	NOT NULL	Código identificador único de la FactTabla de DM_FACT_IDEN_HOGAR_VALIDADO.	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código identificador	Yes

Columna(s) de "DM_FACT_IDEN_HOGAR_VALIDADO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			del Rango de índice de pobreza se enlaza con la dimensión DM_UBIGEO.	
ID_TIEMPO	NUMBER(7)	NOT NULL	Código identificador de Tiempo se enlaza con la dimensión DM_TIEMPO.	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código identificador de Etapa se enlaza con la dimensión DM_ETAPA	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código identificador del Ubigeo se enlaza con la dimensión DM_UBIGEO.	Yes
ID_HOGAR_VALIDADO	NUMBER(7)	NOT NULL	Código identificador del Hogar validado por la comunidad se enlaza con la dimensión DM_HOGAR_VALIDADO.	Yes
N_HOGAR_VALIDADO	NUMBER	NULL	Cantidad de Hogares validados por la comunidad	No

Columna(s) de "DM_FACT_IDEN_POBLACION" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_POBLACION	NUMBER(7)	NOT NULL	Código de identificación único de la población	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO.	Yes
N_POBLACION	NUMBER	NULL	Cantidad de población	No
ID_POBLACION	NUMBER(7)	NOT NULL	Código de población pobre, se enlaza con la dimensión DM_POBLACION	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA.	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código de identificación del umbral, se enlaza con la dimensión DM_UMBRAL_POBREZA.	Yes

Columna(s) de "DM_FACT_IDEN_POBLACION_POBRE" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_POBLACION_POBRE	NUMBER(7)	NOT NULL	Código de identificación único de la población pobre	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código de identificación del umbral, se enlaza con la dimensión DM_UMBRAL_POBREZA.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de	Yes

Columna(s) de "DM_FACT_IDEN_POBLACION_POBRE" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO.	Yes
ID_POBLACION_POBRE	NUMBER(7)	NOT NULL	Código de población pobre, se enlaza con la dimensión DM_POBLACION_POBRE	Yes
N_POBLACION_POBRE	NUMBER	NULL	Cantidad de población pobre	No

Columna(s) de "DM_FACT_IDEN_VIVIENDA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_IDEN_VIVIENDA	NUMBER(7)	NOT NULL	Código de identificación único de la FactTabla.	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código de identificación del umbral, se enlaza con la dimensión DM_UMBRAL.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación del ubigeo se enlaza con la dimensión DM_UBIGEO.	Yes
ID_VIVIENDA	NUMBER(7)	NOT NULL	Código de identificación vivienda, se enlaza con la dimensión DM_VIVIENDA.	Yes
N_VIVIENDA	NUMBER	NULL	Cantidad de viviendas	No

Columna(s) de "DM_FACT_INCO_BENEFICIARIO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_INCO_BENEFICIARIO	NUMBER(7)	NOT NULL	Código de identificación único del beneficiario	Yes
N_BENEFICIARIO	NUMBER	NULL	Cantidad de beneficiarios directos	No
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión	Yes

Columna(s) de "DM_FACT_INCO_BENEFICIARIO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			DM_UBIGEO.	
ID_BENEFICIARIO	NUMBER(7)	NOT NULL	Código de beneficiario, se enlaza con la dimensión DM_BENEFICIARIO	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código de identificación del umbral, se enlaza con la dimensión DM_UMBRAL_POBREZA.	Yes

Columna(s) de "DM_FACT_INCO_CENTRO_EDUCA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_INCO_CENTRO_EDUCA	NUMBER(7)	NOT NULL	Código de identificación único de la FactTabla	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO	Yes
ID_CENTRO_EDUCA	NUMBER(7)	NOT NULL	Código de identificación del umbral, se enlaza con la dimensión DM_CENTRO_EDUCATIVO	Yes
N_CENTRO_EDUCA	NUMBER	NULL	Numero de centros educativos	No

Columna(s) de "DM_FACT_INCO_CENTRO_SALUD" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_INCO_CENTRO_SALUD	NUMBER(7)	NOT NULL	Código de identificación único de la FactTabla	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO	Yes
ID_CENTRO_SALUD	NUMBER(7)	NOT NULL	Código de identificación del centro de salud se enlaza con la dimensión DM_CENTRO_SALUD	Yes
N_CENTRO_SALUD	NUMBER	NULL	Numero de centros de salud	No

Columna(s) de "DM_FACT_INCO_HOGAR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_INCO_HOGAR	NUMBER(7)	NOT NULL	Código identificador único de la FactTabla de DM_FACT_INCO_HOGAR	Yes
ID_UMBRAL	NUMBER(7)	NOT NULL	Código identificador del Rango de índice de pobreza se enlaza con la dimensión DM_UBIGEO.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código identificador de Tiempo se enlaza con la dimensión DM_TIEMPO.	Yes

Columna(s) de "DM_FACT_INCO_HOGAR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_ETAPA	NUMBER(7)	NOT NULL	Código identificador de Etapa se enlaza con la dimensión DM_ETAPA.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código identificador del Ubigeo se enlaza con la dimensión DM_UBIGEO	Yes
N_HOGAR	NUMBER	NULL	Cantidad de Hogares afiliados al PNADP	No
ID_HOGAR	NUMBER(7)	NOT NULL	Código identificador del Hogar afiliado se enlaza con la dimensión DM_HOGAR.	Yes

Columna(s) de "DM_FACT_INCO_PROMOTOR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_INCO_PROMOTOR	NUMBER(7)	NOT NULL	Código de identificación único de la FactTabla.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO	Yes
ID_TIPO_PROMOTOR	NUMBER(7)	NOT NULL	Código de tipo de promotor	Yes
ID_PROMOTOR	NUMBER(7)	NOT NULL	Código de identificación del promotor	Yes
N_PROMOTOR	NUMBER	NULL	Numero de promotores	No

Columna(s) de "DM_FACT_INCO_PROMOTOR_DEP" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_INCO_PROMOTOR_DEP	NUMBER(7)	NOT NULL	Código de identificación único de la FactTabla.	Yes
ID_UBIGEO_DEP	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO a nivel de departamento	Yes
ID_TIPO_PROMOTOR	NUMBER(7)	NOT NULL	Código de tipo de promotor	Yes
ID_PROMOTOR	NUMBER(7)	NOT NULL	Código de identificación del promotor	Yes
N_PROMOTOR	NUMBER	NULL	Numero de promotores	No

Columna(s) de "DM_FACT_LIQ_ABONOS" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_LIQ_ABONOS	NUMBER(7)	NOT NULL	Código de identificación único del abono	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la	Yes

Columna(s) de "DM_FACT_LIQ_ABONOS" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			dimensión DM_UBIGEO	
ID_TITULAR	NUMBER(7)	NOT NULL	Código del Titular, se enlaza con la dimensión DM_TITULAR	Yes
ID_MODALIDAD_PAGO	NUMBER(7)	NOT NULL	Código de identificación de la modalidad de pago, se enlaza con la dimensión DM_MODALIDAD_PAGO.	Yes
N_ABONOS	NUMBER	NULL	Cantidad de abonos realizados.	No
MONTO_ABONOS	NUMBER	NULL	Cantidad en Soles de los abonos realizados.	No

Columna(s) de "DM_FACT_LIQ_CUENTAS" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_LIQ_CUENTAS	NUMBER(7)	NOT NULL	Código de identificación único de la FactTabla.	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA.	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	Yes
ID_CUENTA	NUMBER(7)	NOT NULL	Código de identificación único de la cuenta bancaria, se enlaza con la dimensión DM_CUENTA_BANCARIA.	Yes
N_CUENTAS	NUMBER	NULL	Cantidad de cuentas bancarias	No
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO.	Yes

Columna(s) de "DM_FACT_LIQ_PADRON" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_LIQ_PADRON	NUMBER(7)	NOT NULL	Código de identificación único del Padron	Yes
ID_PADRON	NUMBER(7)	NOT NULL	Código del Padron del PNADP, se enlaza con la dimensión DM_PADRON	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO.	Yes
ID_MODALIDAD_PAGO	NUMBER(7)	NOT NULL	Código de identificación de la modalidad de pago, se enlaza con la dimensión DM_MODALIDAD_PAGO.	Yes
N_ABONOS	NUMBER	NULL	Cantidad de abonos	No

Columna(s) de "DM_FACT_LIQ_PADRON" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			realizados.	
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA.	Yes
MONTO_ABONOS	NUMBER	NULL	Cantidad en Soles de los abonos realizados	No
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	Yes

Columna(s) de "DM_FACT_LIQ_SUSPENDIDOS" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_LIQ_SUSPENDIDOS	NUMBER(7)	NOT NULL	Código de identificación único de la suspensión	Yes
ID_ETAPA	NUMBER(7)	NOT NULL	Código de identificación de la etapa, se enlaza con la dimensión DM_ETAPA	Yes
ID_TIEMPO	NUMBER(7)	NOT NULL	Código de identificación del tiempo, se enlaza con la dimensión DM_TIEMPO.	Yes
ID_UBIGEO	NUMBER(7)	NOT NULL	Código de identificación ubigeo, se enlaza con la dimensión DM_UBIGEO.	Yes
ID_TIPO_SUSPENSION	NUMBER(7)	NOT NULL	Código de identificación del tipo de suspensión, se enlaza con la dimensión DM_TIPO_SUSPENSION.	Yes
ID_SUSPENDIDO	NUMBER(7)	NOT NULL	Código del Titular, se enlaza con la dimensión DM_TITULAR	Yes
ID_MODALIDAD_PAGO	NUMBER(7)	NOT NULL	Código de identificación de la modalidad de pago, se enlaza con la dimensión DM_MODALIDAD_PAGO.	Yes
N_ABONOS	NUMBER	NULL	Cantidad de abonos realizados.	No
MONTO_ABONOS	NUMBER	NULL	Cantidad en Soles de los abonos realizados	No

Columna(s) de "DM_HOGAR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_HOGAR	NUMBER(7)	NOT NULL	Código identificador único del Hogar afiliado.	Yes
X_COMPLEJO	VARCHAR2(32)	NULL	Código identificador único del Hogar en el esquema transaccional del PNADP	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de pobreza del Hogar afiliado.	No

Columna(s) de "DM_HOGAR_ELEGIBLE" Tabla				
Name	Datatype	Null Option	Comment	Is PK

Columna(s) de "DM_HOGAR_ELEGIBLE" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_HOGAR_ELEGIBLE	NUMBER(7)	NOT NULL	Código identificador único del Hogar Elegible.	Yes
X_COMPLEJO	VARCHAR2(32)	NULL	Código identificador único del Hogar en el esquema transaccional del PNADP	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de pobreza del Hogar Elegible	No

Columna(s) de "DM_HOGAR_OBJETIVO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_HOGAR_OBJETIVO	NUMBER(7)	NOT NULL	Código identificador único del Hogar Objetivo	Yes
X_COMPLEJO	VARCHAR2(32)	NULL	Código identificador único del Hogar en el esquema transaccional del PNADP	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de pobreza del Hogar Objetivo	No

Columna(s) de "DM_HOGAR_TOTAL" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_HOGAR_TOTAL	NUMBER(7)	NOT NULL	Código identificador único del Hogar censado.	Yes
X_COMPLEJO	VARCHAR2(32)	NULL	Código identificador único del Hogar en el esquema transaccional del PNADP	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de pobreza del Hogar censado.	No

Columna(s) de "DM_HOGAR_VALIDADO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_HOGAR_VALIDADO	NUMBER(7)	NOT NULL	Código identificador único del Hogar validado por la comunidad.	Yes
X_COMPLEJO	VARCHAR2(32)	NULL	Código identificador único del Hogar en el esquema transaccional del PNADP	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de pobreza del Hogar validado	No

Columna(s) de "DM_MODALIDAD_PAGO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_MODALIDAD_PAGO	NUMBER(7)	NOT NULL	Código Identificador único de la Modalidad de Pago.	Yes
DES_MODALIDAD_PAGO	VARCHAR2(50)	NULL	Descripción de la Modalidad de Pago.	No

Columna(s) de "DM_PADRON" Tabla				
Name	Datatype	Null Option	Comment	Is PK
AGENCIA	VARCHAR2(50)	NULL	Descripción de la Agencia a pagar.	No
ID_PADRON	NUMBER(7)	NOT NULL	Código de identificación único del Padron.	Yes
X_DESESTADO	VARCHAR2(30)	NULL	Almacena el estado del padron: Padron Activo o Padron Pasivo	No
MES_ABONO	VARCHAR2(50)	NULL	mes del abono correspondiente	No
ANO_ABONO	NUMBER(4)	NULL	año del abono	No

Columna(s) de "DM_PADRON" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			correspondiente	
X_UBIGEO	CHAR(6)	NULL	código de Ubigeo	No
MESNUM_ABONO	NUMBER(4)	NULL	numero de mes del abono	No
MONTO	NUMBER(10,2)	NULL	monto a abonar	No
COD_PADRON	CHAR(12)	NULL	Código de identificación único del Padron del esquema transaccional del PNADP.	No
C_TITULAR	CHAR(12)	NULL	Almacena el código de población único del esquema transaccional del PNADP	No
MODALIDAD_PAGO	VARCHAR2(50)	NULL	Descripción de la modalidad de pago el padron	No
NOM_TITULAR	VARCHAR2(150)	NULL	Nombre del titular	No
DNI	CHAR(8)	NULL	dni del titular	No

Columna(s) de "DM_POBLACION" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_POBLACION	NUMBER(7)	NOT NULL	Código de identificación único de la persona	Yes
C_POBLACION	CHAR(12)	NULL	Código de población único del esquema transaccional del PNADP.	No
NOM_POBLACION	VARCHAR2(150)	NULL	Nombre de la persona	No
SEXO_POBLACION	CHAR(10)	NULL	Sexo de la persona	No
FECHA_RECEPCION	DATE	NULL	Fecha de recepción del censo (Fecha de censo)	No
X_DESPOBRE	CHAR(10)	NULL	Descripción del estado de pobreza	No
EDAD_POBLACION	NUMBER	NULL	Edad de la persona	No
X_DESESTUDIA	CHAR(18)	NULL	Descripción de estado estudiante	No
GRADO_INST	VARCHAR2(30)	NULL	Grado de instrucción de la persona	No
X_UBIGEO	CHAR(10)	NULL	Código de ubigeo donde vive la persona	No
X_DESGESTANTE	VARCHAR2(20)	NULL	Descripción del estado gestante	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de Pobreza	No
X_DESVAL_COMUNAL	VARCHAR2(20)	NULL	Descripción de la validación comunal	No

Columna(s) de "DM_POBLACION_POBRE" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_POBLACION_POBRE	NUMBER(7)	NOT NULL	Código de identificación único de la persona pobre	Yes
C_POBLACION	CHAR(12)	NULL	Código de población único del esquema transaccional del PNADP.	No
NOM_POBLACION	VARCHAR2(150)	NULL	Nombre de la persona	No
SEXO_POBLACION	CHAR(10)	NULL	Sexo de la persona	No
EDAD_POBLACION	NUMBER	NULL	Edad de la persona	No
FECHA_RECEPCION	DATE	NULL	Fecha de recepción del censo (Fecha de censo)	No
X_DESGESTANTE	VARCHAR2(20)	NULL	Descripción del estado gestante	No
X_DESVAL_COMUNAL	VARCHAR2(20)	NULL	Descripción de la validación comunal	No
X_DESESTUDIA	CHAR(18)	NULL	Descripción de estado	No

Columna(s) de "DM_POBLACION_POBRE" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			estudiante	
GRADO_INST	VARCHAR2(30)	NULL	Grado de instrucción de la persona	No
X_UBIGEO	CHAR(10)	NULL	Código de ubigeo donde vive la persona	No
X_DESPOBRE	CHAR(10)	NULL	Descripción del estado de pobreza	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de Pobreza	No

Columna(s) de "DM_PROMOTOR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_PROMOTOR	NUMBER(7)	NOT NULL	Código de identificación único del promotor	Yes
DNI_PROMOTOR	CHAR(8)	NULL	DNI del promotor	No
TIPO_PROMOTOR	VARCHAR2(30)	NULL	Tipo de promotor	No
NOM_PROMOTOR	VARCHAR2(100)	NULL	Nombre del promotor	No
DES_CARRERA	VARCHAR2(100)	NULL	Descripción de la carrera del promotor	No

Columna(s) de "DM_RANGO_EDAD" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_RANGO_EDAD	NUMBER(7)	NOT NULL	Código Identificador único del Rango de Edad.	Yes
INI_EDAD	NUMBER	NULL	Edad Inicio del Rango	No
FIN_EDAD	NUMBER	NULL	Edad Fin del Rango.	No
DES_RANGO_EDAD	CHAR(18)	NULL	Descripción del Rango de Edad.	No

Columna(s) de "DM_SUSPENDIDOS" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_SUSPENDIDO	NUMBER(7)	NOT NULL	Código de identificación único de los titulares suspendidos.	Yes
C_TITULAR	CHAR(12)	NULL	Almacena el código de población único del esquema transaccional del PNADP.	No
NOM_TITULAR	VARCHAR2(150)	NULL	Nombre del titular	No
SEXO_TITULAR	CHAR(10)	NULL	Sexo de la persona	No
X_ESTADO_HOGAR	VARCHAR2(30)	NULL	Estado del Hogar del Titular: 1 = Hogar Beneficiario (por lo menos se le ha abonado una vez) y 0 = Hogar Afiliado.	No
FEC_INI_TITULAR	DATE	NULL	Fecha de inicio del Titular, obtenido del esquema transaccional de la tabla titular del campo f_inicio	No
FEC_FIN_TITULAR	DATE	NULL	---Vacio	No
DES_TIPO_SUSPENSION	VARCHAR2(50)	NULL	Descripción del Tipo de Suspensión.	No
COD_TIPO_SUSPENSION	CHAR(2)	NULL	Código del Tipo de Suspensión.	No
FECHA_SUSPENSION	DATE	NULL	Es la fecha que el usuario fue suspendido.	No
X_CONVENIO	CHAR(1)	NULL	Almacena si firmo el convenio con el PNADP: 1 = Firmo, 0 = No Firmo.	No
X_VALIDADO	VARCHAR2(20)	NULL	Almacena el estado de validación comunal del titular, 1= validado por la comunidad, 0 = No	No

Columna(s) de "DM_SUSPENDIDOS" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			validado	

Columna(s) de "DM_TIEMPO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_TIEMPO	NUMBER(7)	NOT NULL	Código Identificador único de Tiempo	Yes
MES_NUM	NUMBER(2)	NULL	Mes número: 1, 2, 3, 4, 11, 12.	No
MES_DES	VARCHAR2(12)	NULL	Mes Descripción: Enero, Febrero, 3, Noviembre, Diciembre.	No
TRIM_NUM	NUMBER(1)	NULL	Trimestre número: 1, 2, 3, 4.	No
TRIM_DES	VARCHAR2(12)	NULL	Trimestre Descripción, ejemplo: Trimestre1.	No
ANO_DES	CHAR(4)	NULL	Año Descripción, ejemplo: '2004'	No
ANO_NUM	NUMBER(4)	NULL	Año número, ejemplo: 2004	No

Columna(s) de "DM_TIPO_PROMOTOR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
COD_TIPO_PROMOTOR	CHAR(3)	NULL	Código del tipo de promotor	No
ID_TIPO_PROMOTOR	NUMBER(7)	NOT NULL	Código de identificación único del tipo de promotor	Yes
DES_TIPO_PROMOTOR	VARCHAR2(50)	NULL	Descripción del tipo de promotor	No

Columna(s) de "DM_TIPO_SUSPENSION" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_TIPO_SUSPENSION	NUMBER(7)	NOT NULL	Código Identificador único del tipo de suspensión.	Yes
COD_TIPO_SUSPENSION	CHAR(2)	NULL	Código del Tipo de Suspensión.	No
DES_TIPO_SUSPENSION	VARCHAR2(50)	NULL	Descripción del Tipo de Suspensión.	No

Columna(s) de "DM_TITULAR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_TITULAR	NUMBER(7)	NOT NULL	Código de identificación único del titular	Yes
C_TITULAR	CHAR(12)	NULL	Almacena el código de población único del esquema transaccional del PNADP.	No
SEXO_TITULAR	CHAR(10)	NULL	Sexo de la persona	No
NRO_CUENTA	CHAR(20)	NULL	Edad de la persona	No
X_ESTADO_HOGAR	VARCHAR2(30)	NULL	Estado del Hogar del Titular: 1 = Hogar Beneficiario (por lo menos se le ha abonado una vez) y 0 = Hogar Afiliado.	No
X_CONVENIO	CHAR(1)	NULL	Almacena si firmo el convenio con el PNADP: 1 = Firmo, 0 = No Firmo.	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de Pobreza	No
X_DESGESTANTE	VARCHAR2(20)	NULL	Descripción del estado gestante	No
X_VALIDADO	CHAR(20)	NULL	Almacena el estado de validación comunal del titular, 1= validado por la comunidad, 0 = No	No

Columna(s) de "DM_TITULAR" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			validado	
X_DESSUSPENDIDO	VARCHAR2(50)	NULL	Descripción de la suspensión del titular	No
FEC_INI_TITULAR	DATE	NULL	Fecha de inicio del Titular, obtenido del esquema transaccional de la tabla titular del campo f_inicio	No
NOM_TITULAR	VARCHAR2(150)	NULL	Nombre del titular	No

Columna(s) de "DM_UBIGEO" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_UBIGEO	NUMBER(7)	NOT NULL	Código identificador único del ubigeo (Lugar geográfico)	Yes
ID_DEP	CHAR(2)	NULL	Código del departamento	No
DES_DEP	VARCHAR2(50)	NULL	Descripción del departamento	No
ID_PROV	CHAR(4)	NULL	Código de la provincia	No
DES_PROV	VARCHAR2(50)	NULL	Descripción de provincia	No
ID_DIST	CHAR(6)	NULL	Código de distrito	No
DES_DIST	VARCHAR2(50)	NULL	Descripción del distrito	No
X_DESCOBERTURA	CHAR(18)	NULL	Descripción de la cobertura del PNADP: Con Cobertura, Sin Cobertura.	No
X_UBIGEO	CHAR(10)	NULL	Código de ubigeo.	No
X_COBERTURA	CHAR(1)	NULL	indicador que señala si el PNADP esta presente en el distrito.	No

Columna(s) de "DM_UBIGEO_DEP" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_UBIGEO_DEP	NUMBER(7)	NOT NULL	Codigo identificador unico de la situacion geografica del departamento	Yes
ID_DEP	CHAR(2)	NOT NULL	Codigo identificador del departamento	No
DES_DEP	VARCHAR2(50)	NOT NULL	Descripcion del departamento	No

Columna(s) de "DM_UMBRAL_POBREZA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_UMBRAL	NUMBER(7)	NOT NULL	Código de identificación único del rango de pobreza	Yes
INI_UMBRAL	NUMBER(10,5)	NULL	Inicio del rango	No
DES_UMBRAL	CHAR(18)	NULL	Descripción del Rango	No
X_DESPOBRE	CHAR(10)	NULL	Descripción del estado del rango: POBRE, NO POBRE	No
FIN_UMBRAL	NUMBER(10,5)	NULL	Final del rango	No

Columna(s) de "DM_VIVIENDA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
ID_VIVIENDA	NUMBER(7)	NOT NULL	Código de identificación único de la vivienda	Yes
COD_VIVIENDA	VARCHAR2(32)	NOT NULL	Código de la vivienda	No
SERVICIO_HIGIENICO	VARCHAR2(50)	NULL	Descripción del servicio higiénico con que cuenta	No
COMBUSTIBLE	VARCHAR2(50)	NULL	Descripcion del combustible utilizado	No

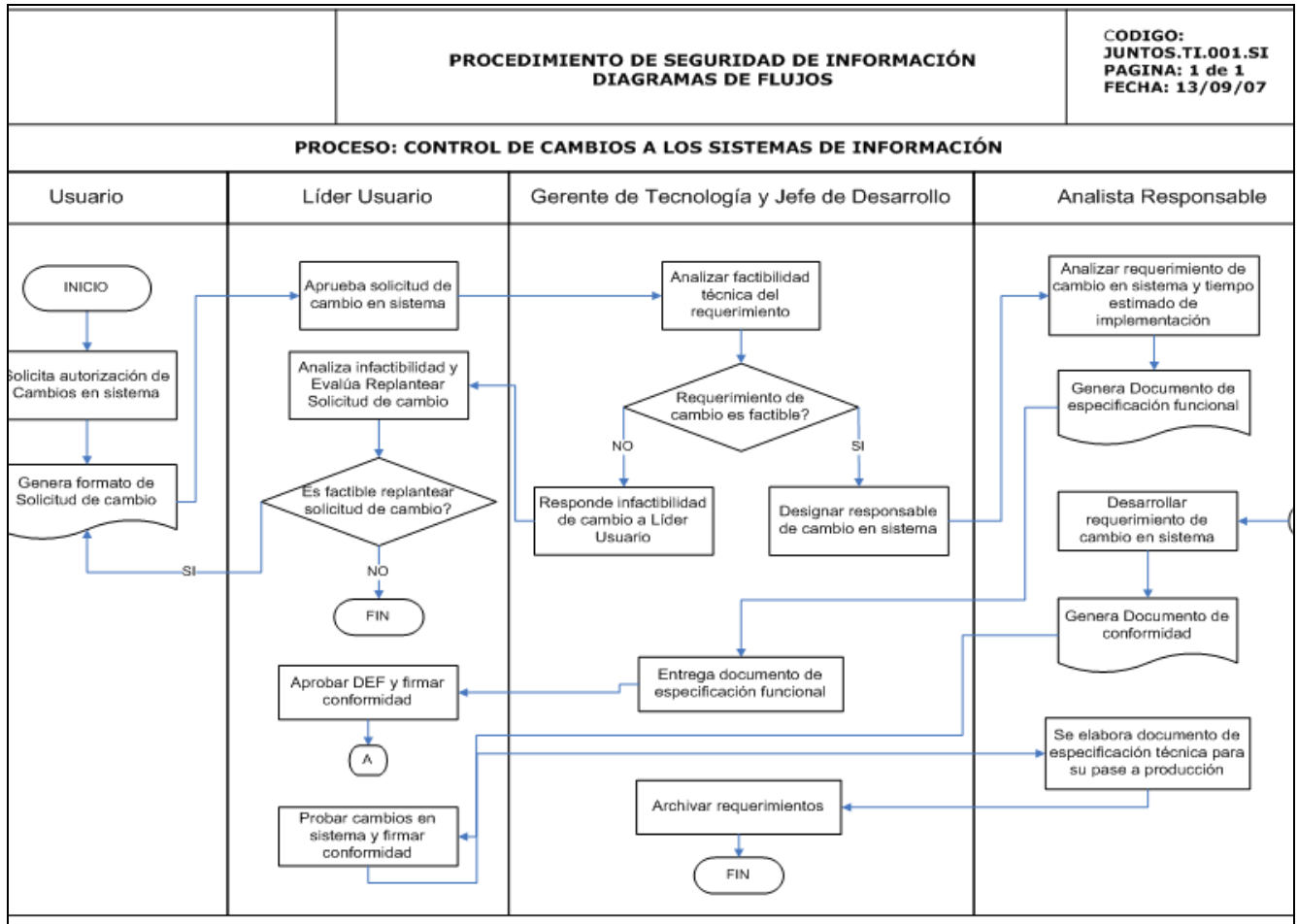
Columna(s) de "DM_VIVIENDA" Tabla				
Name	Datatype	Null Option	Comment	Is PK
			en la vivienda	
X_DESPOBRE	CHAR(10)	NULL	Descripción del estado de pobreza	No
X_UMBRAL	NUMBER(10,5)	NULL	Índice de umbral	No
TIPO_VIVIENDA	VARCHAR2(50)	NULL	Descripción del tipo de vivienda	No
SITUACION_VIVIENDA	VARCHAR2(50)	NULL	Descripción de la situación de la vivienda	No
MATERIAL_PARED	VARCHAR2(50)	NULL	Descripción del material de las paredes de la vivienda	No
MATERIAL_TECHOS	VARCHAR2(50)	NULL	Descripción del material de los techos de la vivienda	No
MATERIAL_PISOS	VARCHAR2(50)	NULL	Descripción del material de los pisos de la vivienda	No
ALUMBRADO	VARCHAR2(50)	NULL	Descripción del alumbrado con que cuenta la vivienda	No
ABASTECE_AGUA	VARCHAR2(50)	NULL	Descripción del tipo de abastecimiento de agua con que cuenta la vivienda	No

ANEXO 2

MAPEO DE DATOS

DM_Ubigeo	ID_UBIGEO	Number(7)	D	JUNTOS	TDEPARTAMENTO	C_DEPA	CHAR(2)	IDENTIFICADOR
DM_Ubigeo	ID_DEP	Char(2)	D	JUNTOS	TDEPARTAMENTO	X_DEPA	VARCHAR2(50)	SELECT
DM_Ubigeo	DES_DEP	Varchar2(50)	D	JUNTOS	TDEPARTAMENTO	C_PROV	CHAR(2)	SELECT
DM_Ubigeo	ID_PROV	Char(4)	D	JUNTOS	TPROVINCIA	X_PROV	VARCHAR2(50)	SELECT
DM_Ubigeo	DES_PROV	Varchar2(50)	D	JUNTOS	TPROVINCIA	C_DIST	CHAR(2)	SELECT
DM_Ubigeo	ID_DIST	Char(6)	D	JUNTOS	TDISTRITO	X_DIST	VARCHAR2(50)	SELECT
DM_Ubigeo	DES_DIST	Varchar2(50)	D	JUNTOS	TDISTRITO	X_DIST	VARCHAR2(50)	SELECT
DM_Ubigeo	X_DESCOBERTURA	varchar2(20)	D	JUNTOS	TDISTRITO	X_UBIGEO	CHAR(6)	SELECT
DM_Ubigeo	X_UBIGEO	Char(10)	D	JUNTOS	TDISTRITO	X_COBERTURA	CHAR(1)	SELECT
DM_Poblacion	ID_POBLACION	Number(7)	D	JUNTOS				IDENTIFICADOR
DM_Poblacion	C_POBLACION	Char(12)	D	JUNTOS	TPOBLACION	X_APPATERNO, X_APMATERNO, X_NOMBRES	CHAR(12)	SELECT
DM_Poblacion	NOM_POBLACION	Varchar2(150)	D	JUNTOS	TPBLACION		VARCHAR(50), VARCHAR(50), VARCHAR(50)	SELECT
DM_Poblacion	SEXO_POBLACION	Char(10)	D	JUNTOS		X_SEXO	CHAR(1)	SELECT
DM_Poblacion	FECHA_RECEPCION	Date	D	JUNTOS	TRECEPCION	F_RECEPCION	DATE	SELECT
DM_Poblacion	X_DESPOBRE	Char(10)	D	JUNTOS	TPOBLACION	X_POBRE	CHAR(1)	SELECT
DM_Poblacion	EDAD_POBLACION	Number	D	JUNTOS	TPOBLACION	N_EDADANO	NUMBER	SELECT
DM_Poblacion	X_DESESTUDIA	Varchar2(20)	D	JUNTOS	TPOBLACION	X_ESTUDIA	CHAR(1)	SELECT
DM_Poblacion	GRADO_INST	Varchar2(30)	D	JUNTOS	TPOBLACION	X_GRADO	CHAR(1)	SELECT
DM_Poblacion	X_UBIGEO	Char(10)	D	JUNTOS	TPOBLACION	C_DEPA, C_PROV, CDIST	CHAR(2), CHAR(2), CHAR(2)	SELECT
DM_Poblacion	X_DESGESTANTE	Varchar2(20)	D	JUNTOS	TPOBLACION	X_GESTANTE	CHAR(1)	SELECT
DM_Poblacion	X_UMBRAL	Number(10,5)	D	JUNTOS	TPOBLACION	X_UMBRAL	NUMBER	SELECT

ANEXO 3



ANEXO 4

